Genomics of neuropeptide signaling system in *Varroa* mites *Varroa destructor* and *Varroa jacobsoni*

by

Rupinder Singh

B.S., Punjab Agricultural University, 2018

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Entomology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2021

Approved by:

Major Professor
Yoonseong Park

Copyright

Abstract

The honey bee is a beneficial insect to humans in various aspects, for example they provide food products like honey, royal jelly, bee wax and propolis used in cosmetic industry as well as pollination services worth $200 billion annually. In the last decade, the global honey bee population has declined at an alarming rate, with a roughly 40% decline in the honey bee population in the USA alone. This decline, commonly called colony collapse disorder (CCD) has been caused by various factors, in particular *Varroa* mite infestations. Various natural and synthetic acaricides were used in the past to control the population of *Varroa* mites. However, the development of acaricide resistance has made it difficult to control the mite infestations over the years. The development of novel chemistries against the *Varroa* mite is not an easy task due to stringent constraints of direct and indirect effects on long-term human health through bee products. One of the possible targets that could be exploited for novel chemistries is the neuropeptide signaling system. Neuropeptides are the signaling molecules for modulating critical physiological processes like growth, development, reproduction, and metabolism. We aimed to identify a specific neuropeptide signaling system present in the *Varroa* mite, but absent in the honey bee using the comparative genomics. Our study focused on conducting the bioinformatics analysis of the neuropeptide signaling system in *Varroa destructor* and *Varroa jacobsoni*. We were able to identify 29 and 31 genes that encode neuropeptides in *V. destructor* and *V. jacobsoni*, respectively. In addition, we identified 27 G protein-coupled receptors that putatively function as neuropeptide receptors. ACP, Allatostatin B, Glycoprotein hormone Beta 5 (GPB5), Proctolin, Natalisin are the neuropeptides for which gene sequences were found in *V. destructor* and *V. jacobsoni*, but were missing in the honey bee.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am especially indebted to my major advisor Dr. Yoonseong Park for believing in my abilities and for his constant support from my first semester here at Kansas State University. I am thankful to him for introducing me to the field of Insect Physiology and Neuroendocrinology that equipped me with the knowledge to implement in this project.

I am also grateful to Dr. Paulina Maldonado for being an amazing mentor during my semesters and throughout my project. I express my gratitude for time and effort spent by Dr. Yoonseong Park and Dr. Paulina Maldonado on providing positive critique and suggestions that molded the outcome of the project.

I would like to express my sincere gratitude to my committee members Dr. Berlin Londono and Dr. Kun Yan Zhu for taking the time to serve on my committee and their support in the process of this project.

I would like to extend a huge thanks to all my family members without whom this journey would not have been possible. And last, but in no way least, I would like to thank my friends for their unwavering support and friendship. The past two years would not have been so memorable and fun without all of them.

Dedication

I dedicate my thesis to my Late Grandfather, S. Kirpal Singh. You will always be in my heart.

Chapter 1 - Introduction

Arthropods

Arthropods are the most diverse group in the animal kingdom, comprising over 1.2 million species (Giribet and Edgecombe, 2019). Members of this group include the arachnids, crustaceans, insects, and myriapods. Arthropods are a very successful animals on earth as seen by their high diversity in habitat distribution (land, water and air), high biomass (80% of all living organisms), and large number in species (Zhang, 2013).

The word arthropod is derived from the Greek root words "arthron" and "podos," meaning joint and foot/leg, respectively, as all arthropods have jointed limbs (Theodor, 1845). Segmentation of the body is seen in arthropods, which helps with classification and distinguishing between various taxa. Insects are identified by their three segmented body; Head, thorax, and abdomen, while crustaceans (e.g. crayfish) or arachnids (e.g. spiders, mites, ticks) have two body segments; cephalothorax and abdomen. A continuous, rigid, chitin, and protein-rich layer called an exoskeleton covers these segments of arthropods as opposed to soft-bodied animals. The exoskeleton protects the arthropod from predation, desiccation, and waterlogging.

Some arthropods cause a significant economic burden as agricultural pests. These pests cause direct damage by spoiling around 18%-26% of global annual crop production, valued at $470 billion (Culliney, 2014). In addition to the direct damage, infested crops also suffer from a loss in quality through reduced nutritional content, reduced market quality, and toxic pest contamination. Other arthropod pests are known for their ability to transmit disease-causing pathogens (such as viruses, bacteria, fungi, nematodes, etc.) to humans and animals. Insects are the vectors for diseases, including but not limited to malaria, chagas disease, yellow fever, and dengue fever. Arachnids are also vectors for diseases including lyme disease, anaplasmosis,

babesiosis, typhus and ehrlichiosis. Apart from diseases, arthropods cause injury via biting, envenomization, and producing allergic reactions.

Aside from the negative impacts caused by some arthropods in agriculture and health, there are many beneficial arthropods that serve essential ecological roles, such as predators of pest species, pollinators, and providers of economically valuable by-products used in different industries. For example, arthropods such as bees, wasps, ants, moths, and butterflies, are primary pollinators for various crops. Predatory arthropods, including the predatory mites and whip scorpion, control the infestations of multiple agricultural pests like aphids, whiteflies and cockroaches, and crickets, respectively. Arthropod-based by-products such as spider webs (spider), and silk (caterpillar) are used to make fishing nets and in the textile industry, respectively (Meyer-Rochow, 2019). Additionally, by-products from bees, such as bee wax and propolis, are used for making candles, furniture wax, polishes, waxed papers, and antiseptics.

Honey bees are one of the most beneficial arthropods to humans. Honey produced by honey bees serves as a food source. Other honey bee by-products are used in various industries. Additionally, honey bees provide economically valuable pollination services for numerous agricultural crops including fruits, nuts, and oilseeds, that make up approximately one-third of the western diet (Brutscher et al., 2016; Klein et al., 2007). The annual estimation of the economic benefits of insect-pollinated crops is evaluated at $175 billion globally, and around $17- 18 billion in countries across North America and European Union (Alló and Loureiro, 2013). To put it in perspective, the world's largest pollination event takes place in the almond fields of California Valley every year in February. Around 1.6 million bees participate in this pollination event. The region alone produces 80% of global almonds, worth $4.8 billion (Brutscher et al., 2016).

During the last two decades, honey bee populations have gradually declined due to factors like ectoparasitic *Varroa* mite infestations, diseases, nutrition, pesticides, and socioeconomic factors. The *Varroa* mite transmits various single-stranded viruses to infested honey bee colonies through their ectoparasitic feeding behavior, causing paralysis in bees which in turn causes population drastically declines. This decline is termed Colony Collapse Disorder (CCD) and currently accounts for 10% honey bee loss (https://beeinformed.org/). According to the National Survey concerning the loss of honey bee populations in the US from 2018 through 2019, there was a reported loss of 38% of managed honey bee colonies during the winter season (1st October 2018 – 1st April 2019). The average rate of population loss increased by 7% from the previous year's rate, marking the highest decline during the winter since the beginning of the survey (in 2006-2007). During the summer season (1st April 2019- 1st October 2019), an estimated 20.5% decline was reported, slightly higher (+3.4 %) than the previous years. In total, there was an average annual estimated loss of 40.7% of commercially managed honey bees during 2018-2019 (https://beeinformed.org/). If the average rate of loss of the honey bee population continues to increase in the future, the world will lose various economically valuable assistances from these beneficial arthropods. The currently used acaricides are ineffective in limiting *Varroa* mites infestations due to the development of resistance in the mites (Hillesheim et al., 1996; Maggi et al., 2009; Milani, 1995 and 1999). Another complicating factor is the lack of knowledge of the biological systems in the *Varroa* mite. Increasing the knowledge of the fundamental biology of the *Varroa* mite will help us in finding a better and efficient control strategy against mite infestations.

Varroa mite

Varroa destructor and *Varroa jacobsoni*

Varroa mite (*V. destructor* and *V. jacobsoni*) belongs to Phylum Arthropoda, Class Arachnida, Subclass Acari, Superorder Parasitiformes. Parasitiformes contain mites that belong to the order Mesostigmata with the extreme diversity of 11,632 described species (Selden, 2016) Mesostigmata is further divided into 12 Suborders, and the *Varroa* mite belongs to Suborder Dermanyssoidea and Family Varroidae. *Varroa* mite has an envelope-like body (without visible segmentation) as a characteristic feature, opposite to what is typically seen in other arthropods. The *Varroa* mite complex contains two different species i.e., *Varroa destructor* and *Varroa jacobsoni*, which act as ectoparasites of honey bees. *V. destructor* parasitizes the Western honey bee *Apis mellifera* while *V. jacobsoni* parasitizes the Asian honey bee *Apis cerana* Fabricius (Anderson and Trueman, 2000).

The *Varroa* mite has three developmental stages: egg, nymph, and adult. The eggs are minute-sized (0.30 mm long and 0.23 mm), white-colored, and oval-shaped. Nymphs have two different forms: protonymph and deutonymph. The protonymph form is white-colored, has pointed mouthparts and eight legs. The protonymph form molts into the deutonymph form, characterized by their reduced setae. A protonymph can be differentiated from a deutonymph based on its circular body shape compared to the oval shape of the deutonymph (Anderson and Trueman, 2000). The life cycle of the mite completes with the final molting from a deutonymph to an adult. The *Varroa* mite (both nymph and adult stage) feeds on the honey bee's fat body for its growth and development (Ramsey et al., 2019). The life cycle of a *Varroa* mite begins with the invasion of the bee brood cell by the adult female *Varroa* mite. The female *Varroa* mite will lay eggs only after the honey bee brood cell has been capped. After 70 hours from the capping of

the bee brood cell, an unfertilized egg is laid, which hatches into a male *Varroa* mite (Huang, 2012). This hatching is based on the haplodiploid sex determination mechanism, where an unfertilized egg hatches into a male offspring, whereas a fertilized egg hatches into a female offspring. The hatched male will then mate with adult female *Varroa* mite, which result in propagation of the mite population in the hive. Female *Varroa* mites lay 5 (in worker/female bee brood cell) – 6 (drone/male bee brood cell) eggs on average. Most males and immature mites fail to emerge from the bee brood cells as these mites die due to dessication in the opened brood cells. Only fully developed and sclerotized adult female mites emerge from bee brood cells. Female mites that have emerged usually go through two phases: the phoretic phase and the reproductive phase. During the phoretic stage, the mites feed on the honey bee's abdominal intersegmental regions by puncturing the soft tissues in this region. The duration of the phoretic stage is determined by the availability of the colony's brood population. It can last anywhere from 5-11 days (when bee brood is available) to 6 months in cold climates (during non-availability of the bee brood). Mites are transferred between bees in the hive during their movement, which spreads the viruses throughout the bee population. Mites can also be transferred to another colony for the following reasons: an infested bee drifts into a nearby colony located near the original colony, infested bees swarm to different locations with other colonies, or a healthy colony transfers to a weak colony.

Varroa mite as the vector of honey bee viruses

Apart from being an ectoparasite, the *Varroa* mite is an efficient vector of bee viruses. Viruses are transmitted within the honey bee population in three stages:

- acquisition of the viral pathogen from infected bee tissues
- movement of the virus-carrying *Varroa* mite among different bees within the colony

- and finally, transmission of the virus into a new healthy bee while feeding on its fat bodies.

Although *Varroa* mite is an efficient vector, secondary conditions such as the presence of pathogen in the mite infested site, the pathogen's survival, the susceptibility of the receiving bees, and whether the mite is a biological vector or a mechanical vector, i.e. whether the virus replicates inside the mite or not, all affect the vector's efficiency (Gisder et al., 2018; Posada-Florez et al., 2019).

The *Varroa* mite is an efficient vector of at least five debilitating bee viruses (Ramsey et al., 2019). Deformed Wing Virus (DWV), Acute Bee Paralysis Virus (ABPV), Kashmir Bee Virus (KBV), Israeli Acute Paralysis Virus (IAPV), and Slow Bee Paralysis Virus (SBPV) are among the viruses vectored by the *Varroa* mite (Moore et al., 2015). The majority of these viruses are single-stranded RNA viruses classified as members of the Family Flaviviridae. Mites transmit viruses throughout the colony when they migrate from one from one bee to another. DWV is highly prevalent in *Varroa* mite infested colonies, with 100% of the bee population infected and virus levels in the bee body being extremely high (de Miranda et al., 2010; Ramsey et al., 2019).

Controlling DWV infection is directly linked to mite parasitism, as horizontal transmission of the virus by the mite increases virulence and subsequent genetic alterations in the virus, which have a significant impact on the bee hive population (Martin et al., 2012; Wilfert et al., 2016). In a colony, adult bees with a DWV infection have deformed wings, shortened abdomens, and discolored cuticle in adult bees. Additionally, bees infected with DWV exhibit altered foraging, aggressiveness (Fujiyuki et al., 2004), and learning behavior (Iqbal and Mueller, 2007). As a result, the colony's strength is significantly reduced, and the colony

eventually collapses. DWV can be transmitted horizontally or vertically. The remaining three viruses, i.e., Acute bee paralysis (ABPV), Kashmir bee virus, and Israel acute paralysis virus (IAPV), form a complex. These viruses are related in terms of the life stages of honey bees they infect and their transmission routes. These viruses infect the honey bee's pupal and adult stages. These viruses are highly virulent because they may rapidly change their titer level in the bee's body from low to high. These are frequently related with colony collapse and can be quite destructive in the colonies afflicted with *Varroa* mite (Genersch, 2010; Genersch et al., 2010).

Current *Varroa* mite management

With honey bees suffering from *Varroa* mite infestation, several control strategies, such as physiological, biological, and chemical are used to eradicate the *Varroa* mite. Physical methods include the use of mite traps, screened bottom boards, and powdered sugar method. Chemical methods employ both synthetic acaricides such as amitraz and coumaphos as well as natural acaricides such as formic acid, oxalic acid and thymol (Adnan et al., 2019). Numerous oils derived from plants such as cinnamon oil, citronella oil, and eucalyptus oil are used to reduce mite infestations (Adnan et al., 2019). However, comprehensive control is not attainable with existing strategies. Currently, chemicals are a major component for limiting *Varroa* mite infestations, however their widespread use has resulted in the development of resistance in the mites. Additionally, honeybee products are contaminated with chemical residues (Adnan et al., 2019). As a result, we must develop alternative effective management measures for the *Varroa* mite in order to address the problem of colony collapse. Alternative control systems could include mite-specific neuropeptide signaling systems that regulate critical physiological processes in the mite. Control strategies targeting the mite-specific neuropeptide signaling

pathway that results in altered physiology and behavior in adult mites could be devised using a biorational design approach.

Neuropeptides in arthropods

Neuropeptides and Neuropeptide repertoire

Neuropeptides are a diverse family of signaling molecules that are involved in the regulation or modulation of arthropod development, growth, reproduction, metabolism, and behavior. These signaling molecules can operate as neurotransmitters in invertebrates in association with the neuromodulation (Nässel, 2018; Nusbaum et al., 2017). In arthropods, molecular signaling involves complex processes supported by many genes that encode peptide precursors (prepropeptides) and receptors (Hauser et al., 2006; Hewes and Taghert, 2001; Nässel and Zandawala, 2019).

A neuropeptide is a signaling molecule that possesses the following characteristics: gene expression, endocrine, and neuronal biosynthesis, storage and controlled release, and ability to regulate functions via interaction with the receptor. Apart from the central nervous system, neuropeptides are also expressed in peripheral tissues, in sensory cells, and in some specific cases, in glial cells, muscle cells, embryonic progenitor cells, and other cells (Nässel and Zandawala, 2019). In general, the molecules are synthesized through transcriptional and translational activation of genes encoding large precursor proteins (prepropeptides or preprohormones), which are then processed at specific cleavage sites to generate shorter and longer peptides. These precursor proteins contain a variety of components, including signal peptides at the precursor’s N-terminus, an active peptide, and mono or dibasic amino acids surrounding mature peptides. Signal peptides are a short chain of amino acids, about 20-25 in number, present at the precursor protein's N-terminus. The signal peptide facilitates the

translocation of precursor proteins into the endoplasmic reticulum's lumen. As the protein is secreted, the signal peptide is cleaved from the propeptide. Prohormone convertase enzymes then process the protein into a mature peptide by utilizing basic amino acids (such as lysine(K) and arginine (R)). Additional posttranslational changes may include amidation at the C-terminus, pyroglutamate cyclization at the N-terminus, and the creation of disulfide bridges. To amidate mature peptides at the C-terminus, the peptidyl aminotransferase (PAM) enzyme utilizes the glycine residue at the C-terminus as an amid donor (Burbach, 2011).

Around 50 distinct neuropeptide genes with unique expression patterns in cells and tissues have been identified in invertebrates. Neuropeptide genes commonly identified in arthropods are: adipokinetic and corazonin like peptide (ACP), adipokinetic hormone (AKH), agatoxin, allatostatins (A, B, C, and CC), allatotropin, arginine vasopressin like, bursicon (alpha and beta), carausious like peptide, cardio acceleratory peptide 2b (CAPA), crustacean cardioactive peptide (CCAP), CCHamide, CNMamide, corazonin, diuretic hormone types (31 and 44), ecdysis triggering hormone (ETH), eclosion hormone, EFLamide, elevenin, FMRFamide, glycoprotein hormones (alpha subunit 2, GPA2 and beta subunit 5, GPB5), gonadulin (insulin like peptide 8), hansolin, insect parathyroid hormone (iPTH), insulin like peptide, ion transport peptide (ITP), limostatin, kinin, myosuppressin (MS), natalisin (NTL), neuropeptide like precursor (NPLP), neuroparsin A, neuropeptide F (NPF), orcokinin, pheromone biosynthesis activating neuropeptide (PBAN), pigment dispersing factor (PDF), proctolin, prothoracicotropic hormone (PPTH), relaxin, RFLamide, RYamide, SIFamide, short neuropeptide F (sNPF), sulfakinin (SK), tachykinin (TK) and trissin.

Neuropeptide receptors

Neuropeptides for modulation of biological functions interact mostly with G Protein-Coupled Receptors (GPCRs, metabotropic), but they can also interact with other receptors such as Tyrosine-Kinase receptors (RTK) (Vogel et al., 2015), Peptide-gated ion channels (FMRFamide-gated sodium channels, ionotropic) (Dürrnagel et al., 2010), and membrane-bound guanylate cyclase receptors (GC) (Chang et al., 2009). GPCRs, RTK, and GC are found in insects, while Peptide gated ion channels are found in Snails and Hydra (Dürrnagel et al., 2010). This study concentrated on GPCRs, the biggest group of neuropeptide receptors found in mammals (Kroeze et al., 2003Z).

G protein-coupled receptors:

GPCRs (seven transmembrane domains) are integral membrane proteins that transduce the extracellular signals (chemicals/ligands) across the membrane via interaction with trimeric G-proteins. These membrane proteins are one of the most numerous and diverse protein families found in mammals (Kroeze et al., 2003Z). The genomes of bacteria, yeast, plants, nematodes, and other invertebrate species all contain the genes encoding GPCRs. GPCRs are classified into five distinct classes: Class A (Rhodopsin type), Class B1 (Secretin), Class B2 (Adhesion type), Class C (Glutamate type), and Class F (Frizzled) (Attwood and Findlay, 1994; Kolakowski Jr, 1994; Hu et al., 2017).

GPCRs regulate a broad range of physiological functions by reacting to and interacting with cellular environmental signals (light, biogenic amines, neuropeptides, peptide hormones, and protons). Normally, G-proteins are present in the trimeric form in the inactive state of the receptor. Once a ligand binds at the receptor site, activation of the receptor occurs through interaction with trimeric G-proteins (Gα, Gβ, and Gγ). Typically, the receptors usually interact

with the Gα subunit through TM3, TM5, TM 6, and intracellular loops 2 and 3 (ICL2 and ICL3) (Hilger et al., 2018). The Gα (alpha subunit) mediates the exchange of cGDP (guanosine diphosphate) with cGTP (guanosine triphosphate) and thereby activates the receptor. When the receptor is active, the Gα dissociates from the Gβγ dimeric subunit. At the downstream level, both Gα and Gβγ regulate the activity of different proteins. The activation of the receptor is completed when cGTP is hydrolyzed back to cGDP, and the Gα reunites with Gβγ dimeric subunit.

GPCRs are potential receptor target sites for discovering novel chemistry/insecticides for use against destructive insect pests. Most of the available insecticidal classes have become ineffective due to resistance development in most insect pests. This interest in arthropod GPCRs is due to the great success of human GPCRs as drug targets. About 35% of pharmaceutical drugs are GPCRs based, while a relatively very low number of chemicals/pesticides on the market are GPCRs based. GPCRs, through their interaction with neuropeptides, regulate various vital physiological processes essential for the survival of the arthropods. Hence, it is proposed that a chemical compound that can disrupt GPCR mediated signaling is a potential candidate for pest control. There is only 1 (formamidine from Class 19) among the 55 different insecticide classes and 30 target sites recognized by the Insecticide Resistance Action Committee (IRAC) that targets GPCRs.

Receptor Tyrosine kinase:

Receptor tyrosine kinases (RTKs) are transmembrane glycoproteins present on the cell surface that have a high affinity for a variety of growth factors, cytokines, and hormones. These transmembrane proteins transfer the extracellular signals (in the form of growth factors) inside the cell via enzymatic phosphorylation (Lemmon and Schlessinger, 2010). The general structure

of an RTK comprises of a variable extracellular ligand-binding domain (a single hydrophobic pass transmembrane helix) and an intracellular protein tyrosine kinase domain (TKD) (ligand-binding domain, or ectodomain). Ligand binding domains or ectodomains interact with a variety of ligands, regulatory cofactors, and other receptors. However, the intracellular part of RTKs vary slightly and commonly contains a single highly conserved TKD. Members of the RTK protein superfamily are involved in a variety of functions, including eukaryotic development, driving cell proliferation, and homeostasis. RTKs, in particular, are critical for controlling cell shape change during migration and morphogenesis, controlling cell and organ growth control, and maintaining and ensuring the survival of both developing and adult tissues.

Neuropeptides, including the ovary edysteroidogenic hormone (OED) and insulin-like peptide (ILP), interact with members of the receptor tyrosine kinase family to regulate a variety of functions by phosphorylating various downstream levels in the insulin signaling pathway (Brown et al., 2008; Dhara et al., 2013). These neuropeptides have been implicated in the regulation of egg formation in *Aedes aegypti* (Vogel et al., 2015). OED is associated with the stimulation of the ovary, which results in uptake of egg yolk by oocytes, whereas ILP is associated with the digestion of the blood meal and the provision of nutrient (like protein) to the egg yolk (Vogel et al., 2015). Ecdysone, the steroid hormone, is associated with stimulation of molting and metamorphosis (sexual maturity) in insects. Prothoracicotropic hormone (PTTH) regulates the production and release of ecdysone (Ou et al., 2016). PTTH modulates this function by interacting with Torso, a tyrosine kinase receptor found in the prothoracic glands of insects. Torso activation via PTTH stimulates the phosphorylation of the cellular signaling molecules associated with the mitogen-activated protein kinase (MAPK) pathway in the Prothoracic gland (Lin and Gu, 2007; Rewitz et al., 2009).

Functions of arthropod neuropeptides

Functional roles of neuropeptides in invertebrates encompass the regulation of homeostasis, organization of behaviors, initiation and coordination of developmental processes, and neuronal and muscular activity's modulation. The number of functions associated with a neuropeptide is species-specific, therefore the following discussion of neuropeptides and their functions will generally hold true for most arthropods but may differ in certain cases.

Modulation of behavior

Neuropeptide-based modulation of the behavior is best described as inducing stereotyped responses as a result of the stimulus induced production of unique motor patterns in neuronal circuits. This modulation is dependent on both the environmental and internal state. Neuropeptides alter the dynamics of neural circuit, hence altering the output motor patterns. Further modulations include changes in the activity of neurons that comprise the circuit or in the synaptic efficiency of the neuronal connections. The below sections describe how neuropeptides influence a variety of invertebrate behaviors, including feeding, reproduction, learning and memory, stress and addiction, circadian rhythm, sleep and wakefulness.

Feeding Behavior:

Feeding behavior involves a series of events that begin with invertebrates experiencing low energy levels, followed by invertebrate seeking food (foraging activity). After finding the food, the invertebrate consumes it until it achieves satiety. The amount of food consumed by an invertebrate is determined by its internal and environmental states, age, and activity. Thus, it indicates the potential involvement of multiple neuropeptides in this behavior. Through neuropeptide injection experiments, one can investigate the effect of neuropeptide signaling on the food intake behavior. Sulfakinin, for example, reduces food consumotion in a variety of

invertebrates, including cockroaches, locusts, bugs, and flies. Sulfakinin's inhibitory effect is the result of its role in stimulating gut contractions (Audsley and Weaver, 2009). Sulfakinin injection significantly reduces food intake in both the fifth instar and adult stages of the locust *Schistocerca gregaria* (Wei et al., 2000; Zels et al., 2015). The effect of sulfakinin signaling is mediated through short Neuropeptide F (sNPF), an anorexic peptide with high levels of expression in the CNS. Numerous observations indicate that sulfakinin receptors and sNPFs are colocalized in the corpora cardiaca of several insects (Predel et al., 2008: Mikani et al., 2012; Dillen et al., 2014; Caers et al., 2015). Injection of the sNPF into silkworms considerably accelerates the onset of feeding (Nagata et al., 2012).

Starvation mediates the increase in reactivity to various stimuli such as odorants, repellents, and pheromones. According to reports, sNPF and CCHamide neuropeptides influence the feeling of starvation (Schoofs et al., 2017). During starvation, flies' antennal lobes exhibit elevated levels of sNPFs and increased expression levels of sNPF receptors. This fact supports the sNPF's potential involvement in food searching behavior (Root et al., 2011). However, some findings indicate the opposite for sNPF. For example, the sNPF injections in *Aedes aegypti* and *Schistocera gregaria* decrease host-seeking behavior (Liesch et al., 2013) and food intake behavior (Dillen et al., 2014).

Allatostatins (AST) and Allatotropin (AT) were initially discovered in cockroaches as a regulator of juvenile hormones released by the corpora allata. These two distinct neuropeptide families are involved in inhibition of food intake behavior. Besides suppressing food intake, AT injections in *Spodoptera frugiperda* results in an increase in larval mortality and a decrease in adult lifespan (Schoofs et al., 2017). Similarly, AST A inhibits food consumption in *Blattella germanica* and starved *D. melanogaster* flies. However, the inhibitory action of AST A is

reduced by the activation of NPF-expressing neurons. Thus, one may assert that many neuropeptides are capable of performing similar functions and are interlinked in their expression and action.

Reproductive behavior:

Similar to feeding behavior, courtship activity is regulated by a variety of neuropeptides in many species. This behavior is very movement intensive and differs between species. For example, courting ritual of the fruit fly consists of a series of steps including chasing, avoidance, dancing, rejection, and copulation (Schoofs et al., 2017). In terms of the specific neuropeptides, the Sex peptide (SP) found in the sperm of male fruit fly modulates the fruit fly's courtship behavior. Male fruit flies leave an antiaphrodiasic pheromone in their sperm after mating, causing the female to repel other approaching males. Following that, females begin laying approximately 80 eggs per day (Kubli and Bopp, 2012).

Natalisin and NPF are two more neuropeptides that are potentially involved in the courtship behavior of fruit flies (Jiang et al., 2013; Kim et al., 2013; Liu et al., 2019). NPF expression in males depends on their courtship state, as it is upregulated in mated males while downregulated in sex-deprived males. Intriguingly, rejected fly males have a higher preference for alcohol, while activation of NPF neurons (representing a mated state) reduces alcohol preference in virgin males. The knockdown of the NPF signaling causes a reduction in male courtship behavior. Additionally, NPF modulates oocyte growth, stimulates vitellogenesis, ecdysteriogenesis, copulation behavior, and fertility (Schoofs et al., 2001; Van Wielendaele et al., 2013).

Stress and Addiction:

Alongside the other behaviors mentioned above, NPF is involved in the regulation of stress and addiction in the fruit fly, *Drosophila melanogaster*. Downregulating NPF expression in the hungry and starved larvae inhibits feeding in the cold climates, while its overexpression causes the fed larvae to feed even on harmful nutrients (Lingo et al., 2007). NPF receptor gene *npfr1*is expressed in the *painless* neurons that participate in the aversion responses to thermal, chemical, and mechanical stressors. In larvae, overexpression of *npfr1 inhibits* sugar-based channel activity, whereas knockdown of the *npfr1* in *painless* neurons results in the functional loss of sugar aversion. These findindgs suggest that the NPF signaling pathway may have potential antinociceptive functions (Xu et al., 2010).

Besides stress regulation in *D. melanogaster*, NPF neurons are linked to ethanol sensitivity. Disruption of the NPF neurons results in resistance to ethanol sedation (Wen et al., 2005). Similarly, corazonin neuropeptide signaling is also associated with ethanol sedation. Adult flies with impaired corazonin signaling have reduced alcohol sensitivity (McClure and Heberlein, 2013). Furthermore, Crz neurons express DH 31 and DH44 receptors. The ligand for these receptors is orthologous to the mammalian stress hormones: calcitonin and corticotropin-releasing hormone. Knockdown of the Crz expressing neurons results in the development of resistance to a variety of stressors such as starvation, high salt concentration, and so on (Zhao et al., 2010). Reduced expression of the DH 44 receptor results in an increase in desiccation tolerance (Cannell et al., 2016).

Regulation of metabolism:

Insect fat bodies are the central storage regions of nutrients obtained from food. These organs serve as a hub for energy storage and utilization through endocrine hormonal interactions

(Arrese and Soulages, 2010). Numerous activities, like insect flight and/or locomotion, require a significant amount of energy. This energy requirement is addressed through the hormonal metabolization of fat bodies. In insects, adipokinetic hormone (AKH) activates catabolic enzymes such as lipases and phosphorylases in insects to accomplish this critical function (Gäde and Auerswald, 2003). For instance, AKH regulates the metabolization of energy reserves in locusts to produce the necessary fuel (carbohydrates, proline, and lipids) for flight (Goldsworthy, 1983; Van der Horst, 2003). Injection of the AKH peptide enhances the lipid concentration in hemolymph in heteropteran species, implying a role in lipid mobilization from fat bodies (Gäde et al., 2006). Similarly, increased glycogen phosphorylase enzyme levels were seen in the hemolymph of *Nezara viridula* following AKH injections (Lorenz and Gäde, 2009).

Modulation of Osmoregulation:

Osmoregulation refers to the regulated recycling of water and ions during solid waste excretion in arthropod species in response to feeding (as in hematophagous insects) and post eclosion. In insects, the osmoregulation is under the control of neural and endocrine hormones called as diuretic and antidiuretic factors (Coast et al., 2002). These hormones are syntehsized in the neurosecretory cells and are stored and released in the corpora cardiaca, and other neuroheamal sites throughout the brain. Tubule secretion regulating hormones in insects are classified into six major families: corticotropin-releasing factor (CRF)-related peptides, calcitonin (CT)-like peptides (CT-DH), Cap_{2b}-like, tachykinin related peptide, insect kinins, and the antidiuretic factors, ADFa and ADFb (Schooley et al., 2012). These hormones exert their action by influencing the activity of two distinct cell types in the Malpighian tubules: principal and stellate cells. Water balance is linked to ion concentration in these cells via activation of enzymes (V-ATPase) and the release of secondary messengers (Ca^{2+}) as a result of diuretic

peptides' engagement with their receptors. For instance, in *D. melanogaster*, binding of the CRF-DH and/or CT-DH ligands to their receptors stimulates the cAMP production, which results in the translocation of protons across the principal cell membrane via the activated apical membrane V-ATPase (Schooley et al., 2012). Cap_{2b} activates the V-ATPase through a different mechanism; it does so through the release of inositol triphosphates (IP_3) as a secondary messenger. IP_3 contributes to the increase of intracellular and extracellular Ca^{2+} levels, hence increasing cGMP levels and activating V-ATPase.

Intriguingly, a peptide can act as both a diuretic and an antidiuretic in different species. For example, Cap_{2b} has diuretic functions in blood-feeding Dipterans such as *A. gambiae* but also anti-diuretic properties in blood-feeding Heteroptera such as *Rhodnius prolixus*. Tenmo-ADFb acts as an antidiuretic peptide in *Tenebrio molitor* but as a diuretic peptide in *Acheta domesticus* (Coast et al., 2007; Schooley et al., 2012). CRF-related DH has been shown to have diuretic activity stimulating Malpighian tubules in a variety of insects, including *A. domesticus* (Coast and Kay, 1994), *Locusta migratoria* (Patel et al., 1995), *D. punctata* (Furuya et al., 2000), *Musa domestica* (Iaboni et al., 1998), and *T. molitor* (Wiehart et al., 2003). CT-DH has been shown to have diuretic properties in insects such as *Diploptera punctata, D. melanogaster, A. gambiae, and R. prolixus.* The CT-like DH peptides often stimulate tubule secretion via the cAMP-dependent pathway, but they can also act through Ca^{2+} signaling pathway. For example, the Drome-DH_{31} or Drome CT-like peptide stimulates the MTs secretion through the cAMP signaling pathway (O'donnell et al., 1996), but the Dippu-DH_{31} and Rhopr-DH_{31} has no effect on cAMP levels or isolated MTs cGMP synthesis. (Coast, 1995; Te Brugge et al., 2005). According to reports, the CT-like peptides in *D. punctata* and *R. prolixus* are reportedly involved in non-selective cation transport, implying that their action is mediated by the Ca^{2+} signaling pathway.

Like Cap_{2b}, insect kinins activate *phospholipase Cβ* in the principal or stellate cells of MTs, as seen in *D. melanogaster* and *A. gambiae*, resulting in an abrupt increase in intracellular Ca^{2+} levels (Radford et al., 2002; Yu and Beyenbach, 2002). Kinins, on the other hand, have a considerable effect on MT secretion in *R. prolixus* (Te Brugge et al., 2002).

Modulation of development:

Insect development includes a series of postembryonic growth cycles which end with the shedding or molting of the hard and rigid exoskeleton through an innate ecdysis motor behavior (Ewer, 2005). This process is governed by a complex interplay between steroid hormones such as the ecdysteroids molting hormones (ecdysone) and the juvenile hormone, as well as neuropeptides/hormones released from the CNS. In insects, the development begins with an increase in ecdysteroid levels in the prothoracic gland as a result of brain-released prothoracicotropic hormone (PTTH). This event is followed by the expression of ETH gene, triggering the production and storage of peptides in the *inka* cells. The level of ecdytseroid influences the competence of the *inka* cells for the release of the peptides (Žitňan et al., 2007; Kingan and Adams, 2000; Ewer, 2005).

Along with the ETH gene, increased ecdysteroid hormone levels activate the two ETH receptors neurons in the CNS, ETHR-A and ETHR-B. These neurons are implicated in the release of peptides from many insects, including eclosion hormone (EH), FMRFamide, kinins, CRF-like diuretic hormones, crustacean cardioactive hormones, myoinhibitory peptide (MIP), and bursicon (Kim et al., 2006a; Kim et al., 2006b). Immunochemical and electrophysiological studies in *Manduca sexta* demonstrate the central release of EH peptides from the ventromedial (VM) cells in the brain in response to the ETH. This action is realted with the synthesis of cGMP in the 27/704 neurons (Hewes and Truman, 1991; Ewer et al., 1997; Gammie and Truman,

1997). Further, the presence of ETHR-A in VM cells suggets that ETH works through EH release and cGMP synthesis (Kim et al., 2006a).

In moths, a network of 27/704 neurons in the abdominal ganglion region produces peptides such as CCAP and MIP. Electrophysiological experiments on isolated CNSs indicate the following: a cocktail of kinins and CRF-like DHs regulates pre-ecdysis, EH is involved in the activation of ETH neurons, and a mixture of CCAP and MIPs regulates the ecdysis motor program (Kim et al., 2006a). Ablated peptidergic neurons result in altered ecdysis behaviors; for example, in *D. melanogaster*, flies with ablated FMRFamide-producing ETHR-A neurons exhibit weak pre ecdysis contractions (Kim et al., 2006b). Similarly, ablation of neurons that produce EH impair tracheal inflation, adult eclosion, and post-eclosion behaviors (McNabb et al., 1997; Clark et al., 2004; Kim et al., 2006b). Furthermore, flies with ablated CCAP neurons exhibit stage-specific abnormalities in ecdysis. These include extended larval pre-ecdysis, substantially impaired pupal ecdysis, and lack of ecdysis and post-ecdysis contractions required for head eversions and appendage extension, all of which result in the death of the majority of flies (Park et al., 2003; Kim et al., 2006b).

Genomics and transcriptomics of the *Varroa* mite

The first attempt at studying the *Varroa* mite's genome occurred in 2002; with the publication of the complete mitochondrial genome sequence of *V. destructor* (NCBI accession # AJ493124) (Navajas et al., 2002). The mitochondrial genome size was approximately 16477 bp in length. Cornman et al., 2010 published a comprehensive genomic study of the *Varroa* mite in 2010, estimating the genome size to be approximately 565 Mbp (Cornman et al., 2010). Homology with eukaryotic genomes was discovered in 13,031 of 184,094 contigs, with an average length of approximately 31.3 Mbp.

The Okinawa Institute of Science and Technology published the current version of the *V. destructor'*s genome Vdes_3.0 at NCBI (accession # GCF 002443255) in 2019 (Techer et al., 2019A). The most recent version of the *V. destructor'*s genome measures 368.9 Mbp in length and contains approximately 30,208 proteins, 12,854 genes, and 107 pseudogenes. In comparison to previous *V. destructor*'s genome assemblies, the current version has 119x greater coverage, higher N50 values for scaffolds and contigs of 58.5 Mbp and 201,886 bp, respectively, and a smaller gap size of 271,335 bp. Among invertebrate genome assemblies, the Vdes_3.0 genome is highly contiguous and has the best-described scaffold N50 values for Acari and Mesostigmata (Techer et al., 2019A).

In the *Varroa jacobsoni*, its whole genome sequence was published recently in 2019 (Techer et al., 2019A) and is available at NCBI available under the name Vjacob_1.0 (accession # GCF_002532875). The genome assembly contains 15,486 genes, 26,243 proteins, and its reported size is 365.59 Mb with 57x coverage, and 408,908 bp gap size. The N50 values for the scaffolds and contigs are 233,810 bp and 96,009 bp, respectively.

These studies involving modern technologies in genome sequencing can provide a much deeper insight into the better understanding of the fundamental biology of the mite species. This valuable information might provide insights into the microbial/parasitic associations or host-pathogen interactions. Comparative genomic analysis of the available genome sequence data can help improve the understanding of the neuropeptide signaling system. Neuropeptide and GPCR gene sequences can be found by conducting a homology-based blast search using their counterpart's gene sequence from the orthologous species. In this study, for conducting the homology-based search, the following four species were used as reference species: *Tetranychus urticae, Ixodes scapularis, D. melanogaster*, and *A. mellifera* (Broeck, 2001; Christie, 2008;

Hauser et al., 2006; Hewes and Taghert, 2001; Veenstra et al., 2012). These species were chosen as reference species based on their phylogenetically close relatedness with *Varroa* mites and the public availability of the fully annotated genome assembly. Also, comparative analysis of the available neuropeptide signaling information in reference species can be utilized to get insights into the evolution of neuropeptide signaling in the *Varroa* mite. There is little information about the neuropeptide genes and receptor genes for *V. destructor* and *V. jacobsoni* available in published articles. The only report about the neuropeptide and neuropeptide signaling in the *Varroa* mite is Tachykinin and Natalisin signaling in *V. destructor* (Jiang et al., 2016).

Direction of the study

Despite the substantial economic impact caused by the *Varroa* mite, little is known about its neuropeptide signaling system. This study aims to find the precursor genes that code for the neuropeptide and their cognate receptors from the genomic data available at NCBI. The study results will help with the understanding of the neuropeptide signaling in *V. destructor* and *V. jacobsoni*. Comparative genomics between the *Varroa* mite and the honey bee will help identify the neuropeptide signal system that could be utilized to design highly selective peptidomimetic acaricides.

Chapter 2 contains an analysis of neuropeptides, including details about the materials and methodologies used to identify the genes, the number of genes identified, information about gene duplication, and concludes with the status of neuropeptides in *V. destructor* and *V. jacobsoni*. Comparison of neuropeptide genes present in *Varroa* mites with other reference species is reported. Suitable figures representing the gene structure of the genes that code for various neuropeptides are included, and detailed insights are discussed.

Chapter 3 contains an analysis of GPCRs, which includes detailed information about the methods and materials employed for finding the potential gene sequences for GPCRs. Results about how many genes are identified, information on gene duplication, and finally, phylogenetic analyses using reference genes are included, and detailed insights are discussed. Chapter 4 contains a discussion and conclusion based on the results from Chapters 2 and 3. The discussion is centerd on the comparative genomic analysis to identify the genes that could be selectively targeted for designing effective control strategies of the *Varroa* mite infestation.

References

1. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., Walter, P., 2013. Molecular biology of the cell. 2015. Garland, New York, 139-194.
2. Alló, M., Loureiro, M.L., 2013. Estimating a meta-damage regression model for large accidental oil spills. Ecological Economics 86, 167-175.
3. Altstein, M., Nässel, D.R., 2010. Neuropeptide signaling in insects, Neuropeptide systems as targets for parasite and pest control. Springer, pp. 155-165.
4. Anderson, D., Trueman, J., 2000. *Varroa jacobsoni* (Acari: Varroidae) is more than one species. Experimental & applied acarology 24, 165-189.
5. Arrese, E.L., Soulages, J.L., 2010. Insect fat body: energy, metabolism, and regulation. Annual review of entomology 55, 207-225.
6. Attwood, T., Findlay, J., 1994. Fingerprinting G-protein-coupled receptors. Protein Engineering, Design and Selection 7, 195-203.
7. Broeck, J.V., 2001. Neuropeptides and their precursors in the fruitfly, Drosophila melanogaster☆. Peptides 22, 241-254.
8. Brutscher, L.M., McMenamin, A.J., Flenniken, M.L., 2016. The buzz about honey bee viruses. PLoS Pathogens 12, e1005757.
9. Chang, J.-C., Yang, R.-B., Adams, M.E., Lu, K.-H., 2009. Receptor guanylyl cyclases in Inka cells targeted by eclosion hormone. Proceedings of the National Academy of Sciences 106, 13371-13376.
10. Christie, A.E., 2008. Neuropeptide discovery in Ixodoidea: an in silico investigation using publicly accessible expressed sequence tags. General and comparative endocrinology 157, 174-185.
11. Clark, A.C., del Campo, M.L., Ewer, J., 2004. Neuroendocrine control of larval ecdysis behavior in Drosophila: complex regulation by partially redundant neuropeptides. Journal of Neuroscience 24, 4283-4292.

12. Cornman, R.S., Schatz, M.C., Johnston, J.S., Chen, Y.-P., Pettis, J., Hunt, G., Bourgeois, L., Elsik, C., Anderson, D., Grozinger, C.M., 2010. Genomic survey of the ectoparasitic mite *Varroa destructor*, a major pest of the honey bee Apis mellifera. BMC genomics 11, 602.
13. Coast, G.M., Orchard, I., Phillips, J.E., Schooley, D.A., 2002. Insect diuretic and antidiuretic hormones.
14. Coast, G.M., 1995. Synergism between diuretic peptides controlling ion and fluid transport in insect Malpighian tubules. Regulatory peptides 57, 283-296.
15. Coast, G.M., Kay, I., 1994. The effects of Acheta diuretic peptide on isolated Malpighian tubules from the house cricket Acheta domesticus. Journal of experimental biology 187, 225-243.
16. Coast, G.M., Nachman, R.J., Schooley, D.A., 2007. An antidiuretic peptide (Tenmo-ADFb) with kinin-like diuretic activity on Malpighian tubules of the house cricket, Acheta domesticus (L.). Journal of Experimental Biology 210, 3979-3989.
17. De Loof, A., Schoofs, L., 1990. Homologies between the amino acid sequences of some vertebrate peptide hormones and peptides isolated from invertebrate sources. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 95, 459-468.
18. de Miranda, J.R., Dainat, B., Locke, B., Cordoni, G., Berthoud, H., Gauthier, L., Neumann, P., Budge, G.E., Ball, B.V., Stoltz, D.B., 2010. Genetic characterization of slow bee paralysis virus of the honey bee (Apis mellifera L.). Journal of General Virology 91, 2524-2530.
19. Dürrnagel, S., Kuhn, A., Tsiairis, C.D., Williamson, M., Kalbacher, H., Grimmelikhuijzen, C.J., Holstein, T.W., Gründer, S., 2010. Three homologous subunits form a high affinity peptide-gated ion channel in Hydra. Journal of Biological Chemistry 285, 11958-11965.
20. Duve, H., Thorpe, A., 1979. Immunofluorescent localization of insulin-like material in the median neurosecretory cells of the blowfly, Calliphora vomitoria (Diptera). Cell and tissue research 200, 187-191.
21. Elphick, M.R., Mirabeau, O., Larhammar, D., 2018. Evolution of neuropeptide signalling systems. Journal of Experimental Biology 221.
22. Ewer, J., 2005. How the ecdysozoan changed its coat. PLoS Biol 3, e349.
23. Ewer, J., Gammie, S.C., Truman, J.W., 1997. Control of insect ecdysis by a positive-feedback endocrine system: roles of eclosion hormone and ecdysis triggering hormone. Journal of Experimental Biology 200, 869-881.
24. Fritsch, H., Van Noorden, S., Pearse, A., 1979. Localization of somatostatin-, substance P- and calcitonin-like immunoreactivity in the neural ganglion of Ciona intestinalis L.(Ascidiaceae). Cell and tissue research 202, 263-274.
25. Furuya, K., Milchak, R.J., Schegg, K.M., Zhang, J., Tobe, S.S., Coast, G.M., Schooley, D.A., 2000. Cockroach diuretic hormones: characterization of a calcitonin-like peptide in insects. Proceedings of the National Academy of Sciences 97, 6469-6474.
26. Fujiyuki, T., Takeuchi, H., Ono, M., Ohka, S., Sasaki, T., Nomoto, A., Kubo, T., 2004. Novel insect picorna-like virus identified in the brains of aggressive worker honey bees. Journal of virology 78, 1093-1100.
27. Gäde, G., Auerswald, L., Marco, H.G., 2006. Flight fuel and neuropeptidergic control of fuel mobilisation in the twig wilter, Holopterna alata (Hemiptera, Coreidae). Journal of insect physiology 52, 1171-1181.
28. Gammie, S.C., Truman, J.W., 1997. Neuropeptide hierarchies and the activation of sequential motor behaviors in the hawkmoth, Manduca sexta. Journal of Neuroscience 17, 4389-4397.

29. Genersch, E., 2010. Honey bee pathology: current threats to honey bees and beekeeping. Applied microbiology and biotechnology 87, 87-97.
30. Genersch, E., Von Der Ohe, W., Kaatz, H., Schroeder, A., Otten, C., Büchler, R., Berg, S., Ritter, W., Mühlen, W., Gisder, S., 2010. The German bee monitoring project: a long term study to understand periodically high winter losses of honey bee colonies. Apidologie 41, 332-352.
31. Gisder, S., Möckel, N., Eisenhardt, D., Genersch, E., 2018. In vivo evolution of viral virulence: switching of deformed wing virus between hosts results in virulence changes and sequence shifts. Environmental microbiology 20, 4612-4628.
32. Goldsworthy, G., 1983. The endocrine control of flight metabolism in locusts. Advances in insect physiology 17, 149-204.
33. Grimmelikhuijzen, C.J., Williamson, M., Hansen, G.N., 2004. Neuropeptides in cnidarians, Cell signalling in prokaryotes and lower Metazoa. Springer, pp. 115-139.
34. Hauser, F., Cazzamali, G., Williamson, M., Blenau, W., Grimmelikhuijzen, C.J., 2006. A review of neurohormone GPCRs present in the fruitfly Drosophila melanogaster and the honey bee Apis mellifera. Progress in neurobiology 80, 1-19.
35. Hewes, R.S., Truman, J.W., 1991. The roles of central and peripheral eclosion hormone release in the control of ecdysis behavior in Manduca sexta. Journal of Comparative Physiology A 168, 697-707.
36. Hewes, R.S., Taghert, P.H., 2001. Neuropeptides and neuropeptide receptors in the Drosophila melanogaster genome. Genome research 11, 1126-1142.
37. Hilger, D., Masureel, M., Kobilka, B.K., 2018. Structure and dynamics of GPCR signaling complexes. Nature structural & molecular biology 25, 4.
38. Hu, G.-M., Mai, T.-L., Chen, C.-M., 2017. Visualizing the GPCR network: classification and evolution. Scientific reports 7, 1-15.
39. Huang, Z., 2012. *Varroa* mite reproductive biology. American Bee Journal, Bee Culture (Joint publication), http://www. extension. org/pages/65450/*Varroa*-mite-reproductive-biology.
40. Iaboni, A., Holman, G.M., Nachman, R.J., Orchard, I., Coast, G.M., 1998. Immunocytochemical localisation and biological activity of diuretic peptides in the housefly, Musca domestica. Cell and tissue research 294, 549-560.
41. Iqbal, J., Mueller, U., 2007. Virus infection causes specific learning deficits in honey bee foragers. Proceedings of the Royal Society B: Biological Sciences 274, 1517-1521.
42. Jiang, H., Kim, D., Dobesh, S., Evans, J.D., Nachman, R.J., Kaczmarek, K., Zabrocki, J., Park, Y., 2016. Ligand selectivity in tachykinin and natalisin neuropeptidergic systems of the honey bee parasitic mite *Varroa destructor*. Scientific reports 6, 19547.
43. Kim, Y.-J., Žitňan, D., Cho, K.-H., Schooley, D.A., Mizoguchi, A., Adams, M.E., 2006a. Central peptidergic ensembles associated with organization of an innate behavior. Proceedings of the National Academy of Sciences 103, 14211-14216.
44. Kim, Y.-J., Žitňan, D., Galizia, C.G., Cho, K.-H., Adams, M.E., 2006b. A command chemical triggers an innate behavior by sequential activation of multiple peptidergic ensembles. Current Biology 16, 1395-1407.
45. Kingan, T.G., Adams, M.E., 2000. Ecdysteroids regulate secretory competence in Inka cells. Journal of Experimental Biology 203, 3011-3018.

46. Klein, A.-M., Vaissiere, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T., 2007. Importance of pollinators in changing landscapes for world crops. Proceedings of the royal society B: biological sciences 274, 303-313.
47. Kolakowski Jr, L.F., 1994. GCRDb: a G-protein-coupled receptor database. Receptors & channels 2, 1-7.
48. Li, B., Predel, R., Neupert, S., Hauser, F., Tanaka, Y., Cazzamali, G., Williamson, M., Arakane, Y., Verleyen, P., Schoofs, L., 2008. Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle Tribolium castaneum. Genome research 18, 113-122.
49. Liu, F., Baggerman, G., Schoofs, L., Wets, G., 2006. Uncovering conserved patterns in bioactive peptides in Metazoa. Peptides 27, 3137-3153.
50. Lorenz, M.W., Gäde, G., 2009. Hormonal regulation of energy metabolism in insects as a driving force for performance. Integrative and Comparative Biology 49, 380-392.
51. Martin, S.J., Highfield, A.C., Brettell, L., Villalobos, E.M., Budge, G.E., Powell, M., Nikaido, S., Schroeder, D.C., 2012. Global honey bee viral landscape altered by a parasitic mite. Science 336, 1304-1306.
52. McNabb, S.L., Baker, J.D., Agapite, J., Steller, H., Riddiford, L.M., Truman, J.W., 1997. Disruption of a behavioral sequence by targeted death of peptidergic neurons in Drosophila. Neuron 19, 813-823.
53. Mirabeau, O., Joly, J.-S., 2013. Molecular evolution of peptidergic signaling systems in bilaterians. Proceedings of the national academy of sciences 110, E2028-E2037.
54. Moore, P.A., Wilson, M.E., Skinner, J.A., 2015. Honey bee viruses, the deadly *Varroa* mite associates. Bee Health 19, 2015.
55. Navajas, M., Conte, Y.L., Solignac, M., Cros-Arteil, S., Cornuet, J.-M., 2002. The complete sequence of the mitochondrial genome of the honey bee ectoparasite mite *Varroa destructor* (Acari: Mesostigmata). Molecular biology and evolution 19, 2313-2317.
56. O'donnell, M., Dow, J., Huesmann, G., Tublitz, N., Maddrell, S., 1996. Separate control of anion and cation transport in malpighian tubules of Drosophila Melanogaster. Journal of Experimental Biology 199, 1163-1175.
57. Ou, Q., Zeng, J., Yamanaka, N., Brakken-Thal, C., O'Connor, M.B., King-Jones, K., 2016. The insect prothoracic gland as a model for steroid hormone biosynthesis and regulation. Cell reports 16, 247-262.
58. Park, J.H., Schroeder, A.J., Helfrich-Förster, C., Jackson, F.R., Ewer, J., 2003. Targeted ablation of CCAP neuropeptide-containing neurons of Drosophila causes specific defects in execution and circadian timing of ecdysis behavior. Development 130, 2645-2656.
59. Patel, M., Hayes, T., Coast, G., 1995. Evidence for the hormonal function of a CRF-related diuretic peptide (Locusta-DP) in Locusta migratoria. Journal of experimental biology 198, 793-804.
60. Posada-Florez, F., Childers, A.K., Heerman, M.C., Egekwu, N.I., Cook, S.C., Chen, Y., Evans, J.D., Ryabov, E.V., 2019. Deformed wing virus type A, a major honey bee pathogen, is vectored by the mite Varroa destructor in a non-propagative manner. Scientific reports 9, 1-10.
61. Radford, J.C., Davies, S.A., Dow, J.A., 2002. Systematic G-protein-coupled receptor analysis inDrosophila melanogaster identifies a leucokinin receptor with novel roles. Journal of Biological Chemistry 277, 38810-38817.

62. Ramsey, S.D., Ochoa, R., Bauchan, G., Gulbronson, C., Mowery, J.D., Cohen, A., Lim, D., Joklik, J., Cicero, J.M., Ellis, J.D., 2019. *Varroa destructor* feeds primarily on honey bee fat body tissue and not hemolymph. Proceedings of the National Academy of Sciences 116, 1792-1801.
63. Schooley, D.A., Horodyski, F.M., Coast, G.M., 2012. Hormones controlling homeostasis in insects, Insect endocrinology. Elsevier, pp. 366-429.
64. Selden, P.A., 2016. Arachnida, Las Hoyas: A Cretaceous Wetland.: A multidisciplinary synthesis after 25 years of research on an exceptional fossil Lagerstätte from Spain. Friedrich Verlag, pp. 64-67.
65. Te Brugge, V., Lombardi, V., Schooley, D., Orchard, I., 2005. Presence and activity of a Dippu-DH31-like peptide in the blood-feeding bug, Rhodnius prolixus. Peptides 26, 29-42.
66. Thorndyke, M., Probert, L., 1979. Calcitonin-like cells in the pharynx of the ascidian Styela clava. Cell and tissue research 203, 301-309.
67. Van der Horst, D.J., 2003. Insect adipokinetic hormones: release and integration of flight energy metabolism. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 136, 217-226.
68. Veenstra, J.A., Rombauts, S., Grbić, M., 2012. In silico cloning of genes encoding neuropeptides, neurohormones and their putative G-protein coupled receptors in a spider mite. Insect biochemistry and molecular biology 42, 277-295.
69. Wiehart, U., Nicolson, S., Van Kerkhove, E., 2003. K+ transport in Malpighian tubules of Tenebrio molitor L.: a study of electrochemical gradients and basal K+ uptake mechanisms. Journal of experimental biology 206, 949-957.
70. Wilfert, L., Long, G., Leggett, H., Schmid-Hempel, P., Butlin, R., Martin, S., Boots, M., 2016. Deformed wing virus is a recent global epidemic in honeybees driven by Varroa mites. Science 351, 594-597.
71. Yu, M.-J., Beyenbach, K.W., 2002. Leucokinin activates Ca2+-dependent signal pathway in principal cells of Aedes aegypti Malpighian tubules. American Journal of Physiology-Renal Physiology 283, F499-F508.
72. Žitňan, D., Kim, Y.-J., Žitňanová, I., Roller, L., Adams, M., 2007. Complex steroid–peptide–receptor cascade controls insect ecdysis. General and comparative endocrinology 153, 88-96.

Chapter 2 - Annotation of *Varroa* mite neuropeptides

Introduction

Neuropeptides are small proteins that regulate arthropod development, growth, reproduction, metabolism, and behavior (Nässel, 2018; Nusbaum et al., 2017). Many genes that encode peptide precursors (prepropeptides) and their receptors are involved in peptidergic signaling (Hauser et al., 2006; Hewes and Taghert, 2001; Nässel and Zandawala, 2019). Neuropeptides are mainly produced in the central nervous system (neurons and neuroendocrine cells). Their storage and release are regulated in response to temporal and spatial conditions (Bendena, 2010). Neuropeptides, in addition to the central nervous system, are also found in peripheral tissues, endocrine cells, sensory cells, and in some specific cases, glial cells, muscle cells, embryonic progenitor cells (Nässel and Zandawala, 2019).

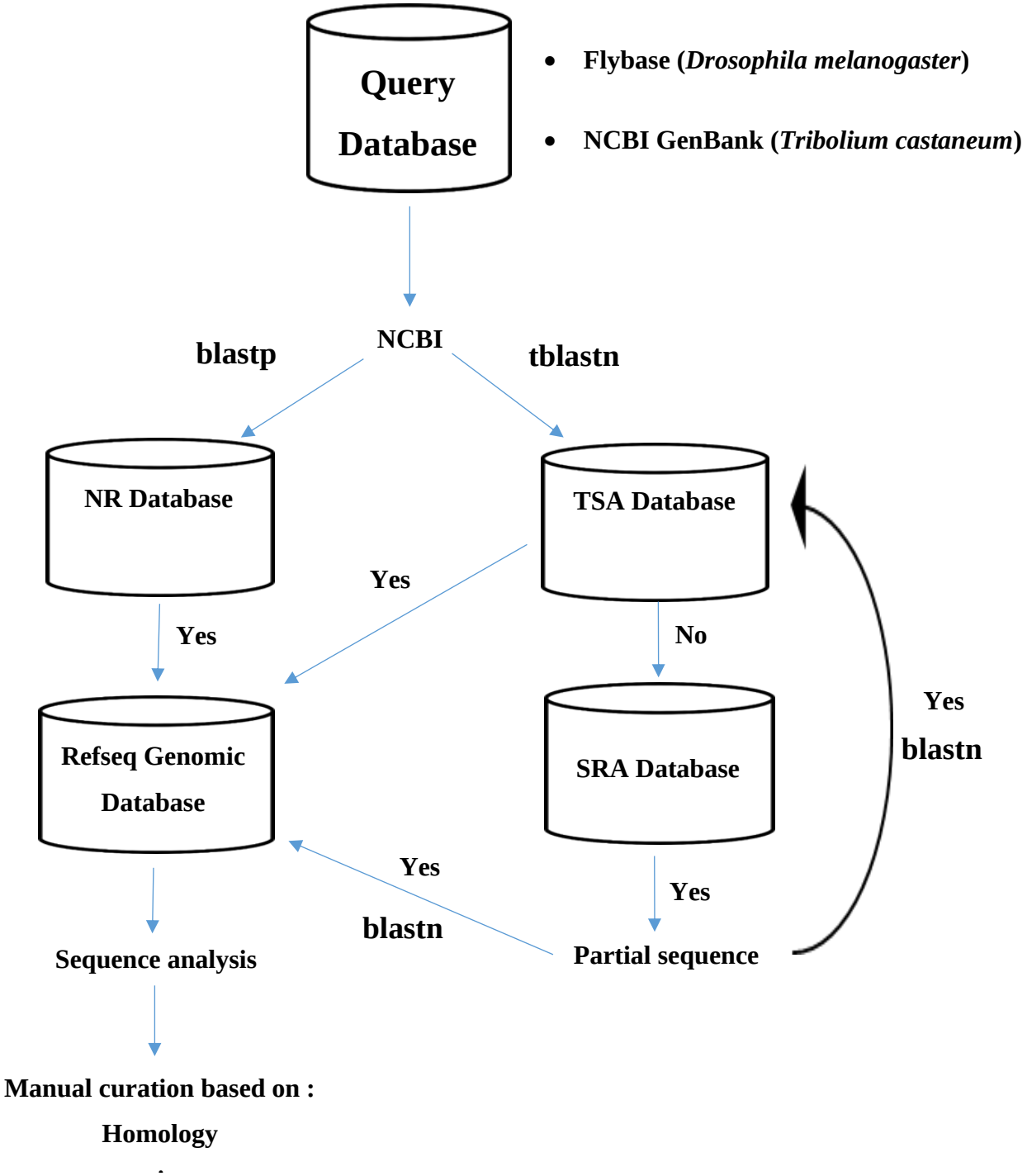
The synthesis of neuropeptide molecules begins with the transcriptional activation of genes, with their translation producing precursor proteins (prepropeptide or preprohormone). The precursor proteins are composed of several components which includes the signal peptide at the precursor's N-terminus, active peptides and canonical mono or dibasic amino acids surrounding mature peptides. Active peptides are often formed through posttranslational modifications and processing at specific cleavage sites. Signal peptides are a short chain of 20 to 25 amino acids located at the N-terminus of the precursor protein (Bendena, 2010; Kapp et al., 2009). The signal peptide facilitates the translocation of precursor proteins into the lumen of the endoplasmic reticulum. The propeptide is further processed posttranslationally in the endoplasmic reticulum-golgi network (Bendena, 2010; Li and Kim, 2008). Prohormone convertase enzymes utilize the essential amino acids (like lysine, K, and arginine, R) to process mature peptides. Additional posttranslational modifications may include carboxy (C) terminal amidation, N-terminal

pyroglutamate cyclization, and the formation of intramolecular and intermolecular disulfide bonds. For C-terminal amidation of mature peptides, peptidylglycine α-amidating monooxygenase (PAM) related enzymes such as peptidylglycine α-hydroxylating monooxygenase (PHM), and peptidyl α-hydroxyglycine-α-amidating lyase (PAL), utilize the glycine residue present at C terminal end, as an amide donor (Kolhekar et al., 1997).

In modern biology, bioinformatics has become an important tool for providing valuable information that gives insight into the evolution of neuropeptide signaling across various taxa. In the Phylum Arthropoda, around 30–50 different neuropeptide genes with unique expression patterns in cells and tissues have been identified in each species. For instance, 35 neuropeptide genes were found in the malarial mosquito, *Anopheles gambiae* (Riehle et al., 2002), while 36 were found in the honey bee, *Apis mellifera* (Hummon et al., 2006), respectively. Expanded genome analysis, with the depth for listing the complete sets of neuropeptide repertoires, provides an understanding of the evolutionary patterns of the genes and their functional insights.

Before this study, there was no publicly available information on the neuropeptide signaling of the *Varroa* mite. The objective of this study was to identify neuropeptide precursor genes from the *Varroa* mite's genomic scaffold (*Varroa destructor* and *Varroa jacobsoni*). A better understanding of the *Varroa* mite's neuropeptide repertoire and how the neuropeptide signaling has evolved in arachnids will be achieved by comparing it with species from closely related taxa, including Class Insecta, *D. melanogaster*, and *A. mellifera*, and Class Chelicerata, *Ixodes scapularis*, and *Tetranychus urticae*. Additionally, the findings of the study will aid in identifying a neuropeptide that could be used to develop bee safe acaricides.

Figure 2.1: Schematic flow of homology blast search



Initially, the search of the sequences that encode preprohormones were made in the non-redundant database (nr) for the *Varroa* mite at the National Centre for Biotechnology Information (NCBI, https://blast.ncbi.nlm.nih.gov/Blast.cgi), which mainly contains reference annotated sequences. Using a variety of search parameters, the blastp program was employed to search for protein sequences in *Drosophila melanogaster*, *Apis mellifera*, *Ixodes scapularis*, and *Tetranychus urticae*. The parameter matrix used was either Blosum 62 or PAM 30, and the expected threshold value was in the range of 10 to 1000. The manual correction of the nr data in NCBI was made based on the conserved sequence motif of the active peptides. Further annotations were made to the putative coding regions (CDS) for identifying the post-translational processing sites.

If the search in the nr database did not yield a significant matching sequence, the search was then expanded to other databases at NCBI; RefSeq genome database, Transcriptome shotgun assembly database (TSA), Sequence read archive database (SRA)(accession numbers for *V. destructor*: SRX5654493, SRX5651414, SRX4926906, SRX2960666, SRX8188313, SRX2960669, SRX1955933, and for *V. jacobsoni: SRX1825493, SRX1825428, SRX1825392, SRX1825391, SRX1825346, SRX1824269)* using the tBlastn program. The search strategy was the same as the above descriptions. The match found in the RefSeq genome database was further processed for obtaining full open reading frame by using the FGENESH program (http://www.softberry.com/) with the input as the sequence surrounding the matching region; 1500 - 2000 kb of each side of the match was entered in to the program to find the potential entire coding sequence. In the case of the potential sequences matching to the sequence in SRA database for mRNA sequences, the complete gene sequence was identified by conducting repeated blast searches for walking towards 5’ and 3’ directions to find the full-length mRNA

sequence. The searches for mRNA walking usually extended the range of the transcript but did not reach the full open reading frame. The presence of the putative mature peptide in the sequence was the main criteria for considering it as a part of neuropeptide transcript. The partial sequence obtained from SRA data walking was used in a blast searches in the TSA database and Refseq genomic database using the partial gene sequence as a query. If the query failed to find a potential match in either of the databases, the sequence was considered as a putative partial neuropeptide sequence.

After identifying the putative sequences encoding the neuropeptide, further analysis was made to annotate signal peptides, canonical monobasic or dibasic cleavage sites, and mature peptides. The signal peptide found at the N-terminus of the sequence was predicted using SignalP 5.0 server (Armenteros, Tsirigos et al. 2019) (http://www.cbs.dtu.dk/services/SignalP/). Characterization of the basic amino acid as a site for proteolytic processing was based on two factors: the presence or absence of basic (Lysine, Arginine) or aliphatic amino acids and their position relative to the putative site (Veenstra 2000). Comparison of the mature peptide with their counterparts in other ortholgous genes from closely related species was made by aligning their sequences using the Muscle alignment algorithm (Edgar 2004).

Results

In the current study, a total of 32 neuropeptide gene sequences (6 partial and 26 full length) were identified in the genomic scaffold of *V. destructor* and 34 neuropeptide gene sequences (7 partial and 27 full length) were identified in the genomic scaffold of *V. jacobsoni*. Out of the 32 genes identified in *V. destructor*, 18 were found from nr, 5 from TSA, and 8 genes from SRA databases. Out of the 34 genes identified from *V. jacobsoni*, 17 were found from nr, 9 from TSA, and 7 genes from SRA databases. In *V. destructor*, the genes are predicted to code for

47 mature peptides, while in *V. jacobsoni*, the number reaches to 49 mature peptides. Table 1 shows the summary of the search results, excluding the 18 genes sequences which were not found in the genomic database of *Varroa* mites that encode for: AKH, allatostatin CC, allatotropin, corticotropic releasing factor-like DH (37/47), eclosion hormone, ecdysis triggering hormone, glycoprotein hormone alpha-2, gonadulin, leucokinin, neuroparsin A, orcokinin B, prothoracicotropic hormone, RYamide, and short neuropeptide F (sNPF). In the case of myosuppressin and orcokinin B gene, their genes were identified only in *V. jacobsoni* but not in *V. destructor*.

Conservation in orthologs:

The neuropeptide genes identified in this study have conserved sequence motifs found in various species belonging to Phylum Arthropoda, indicating evolutionary conservation of the gene. The conservation in the sequence of neuropeptide genes can be described in various aspects, such as the presence of conserved sequence motifs, multiple bioactive peptides/mature peptides in the similar precursor sequence, and gene structures with conserved locations of the mature peptides in the precursor sequence of the gene.

Conserved sequence motifs for identifying neuropeptides:

The orthologues could be identified with the help of the conserved sequence motif, i.e., the occurrence of an amino acid at a position relative to the C or N- termini regions of the mature peptide. For example, Allatsotatin A can be identified with the motif FGL/I (Phe-Gly-Leu/Ile) as in *A. mellifera* and *T. urticae*, Allatostatin B with the motif WX6W-amide (Trp-X6-Trp-amide), Proctolin with the motif RYLPT (Arg-Tyr-Leu-Pro-Thr), SIFamide with the motif SIF-amide (Ser-Ile-Phe-amide), and Myosuppressin with the motif X1DVX2HX3FLRF-amide (X-Asp-X-X-His-X-Phe-Leu-Arg-Phe-amide), located at the C-terminal of the mature peptide among

various species in arthropods (Egerod et al., 2003). Similarly, other neuropeptide genes identified in this study also have their conserved sequence motif in their mature peptides. Furthermore, a comparison of mature peptide sequences among various arthropods, using multiple sequence alignment, supports the conserved nature of neuropeptide genes.

Tandem repeats of mature peptides in single genes:

In addition to the conservation of sequence motifs, repeat patterns of mature peptides were also found in some neuropeptide gene sequences. Mono or dibasic amino acids surround the mature peptide to help endonuclease restriction at these sites. Thus, the mature peptides are cleaved from the precursor gene sequence during posttranslational modifications. The repeats of mature peptides within the precursor were seen in orthologous genes in different species with a varied number of repeats in a species-specific manner. For example, the precursor gene sequence of allatostatin B found in the *Varroa* mite contains three repeats, while the precursor gene sequence of allatostatin B found in *Drosophila melanogaster* contains five repeats. In Class Insecta/Hexapoda, the gene sequence encoding allatostatin B in *D. melanogaster* has five mature peptide repeats; in Class Crustacea, the precursor gene sequence of allatostatin B in *Nephrops norvegicus* has five mature peptide repeats, while the same sequence in *Daphnia pulex* has seven; and in Class Chelicerata, the precursor gene sequence of allatostatin B in *Limulus polyphemus*, *Galendromus occidentalis*, and *T. urticae* have nine, three, and six mature peptide repeats, respectively. Similarly, other genes like Ast A, CAPA, FMRFamide, etc., found in *Varroa* mites also have multiple mature peptide repeats (two, eight, and eleven respectively) separated by dibasic or monobasic cleavage sites (Supplementary data).

Location of mature peptide:

The location of the mature peptide in the preprohormone gene sequence is also found to be conserved. For example, the mature peptide sequence of proctolin neuropeptide is located immediately after the signal peptide. It is followed by a canonical dibasic or monobasic amino acid, which act as an endonuclease restriction site for separation from associated proteins that are about 100 amino acid long.

Unusual variations in the conserved amino acid motif in FMRFamide:

While the mature peptide sequence of the FMRFamide neuropeptide gene is mostly conserved across the Phylum Arthropoda, the mature peptide sequence of the FMRFamide neuropeptide gene identified from the *Varroa* mite is slightly different. Rather than the typical FMRF sequence motif at the C terminus, the gene contains an FMHF sequence motif. An FMRFamide neuropeptide is a collection of related peptides that share the FaRF consensus sequence motif, (Phe-Ile/Val/Met/Leu-Arg-Phe-amide); other variants include FIRFamide (Phe-Ile-Arg-Phe-amide) and IVRFamide (Ile-Val-Arg-Phe-amide). In 1977, the FMRFamide neuropeptide was isolated from the clam *Mercenaria mercenaria* (Price and Greenberg, 1977). Multiple copies of FMRFamide-related peptides have been identified in the precursor genes in arthropods dating all the way back to 1990, where the discovery FMRFamide peptide in the form of FMHFamide was first reported from the *D. melanogaster* (Taghert and Schneider, 1990). FMHFamide, FIRFamide, and FVRSamide were all isolated from *Mercenaria mercenaria*. Similarly, multiple FMRFamide peptides are typically present in orthologous counterparts from various species. Comparing at the Phylum level, only *D. melanogaster* and *D. virilis* (Class Insecta) possess mature peptides with the FMHF sequence rather than the FMRF sequence. Along with the *Varroa* mite, the FaHF sequence motif in FMRFamide peptide is found in the predatory mite *Glanderomus occidentalis* (Class Chelicerata). Simultaneously, the remainder of

the members of the same class have FMRF as a C-terminus sequence motif. It is possible that switching from arginine to positively charged histidine retains the receptor's activity, depending on the species.

Novel neuropeptide genes:

Recent insect genome analysis reports indicate an increased neuropeptide repertoire with the findings of new neuropeptide genes, including carausious-like neuropeptide and agatoxin-like neuropeptide. The carausious-like neuropeptide was reported from the stick insect, *Carausius morosus* (Liessem et al., 2018), and the American cockroach, *Periplaneta americanum* (Zeng et al., 2020). Through the blast search, carausious-like neuropeptide gene was identified with a percent identity of 43.48% and an e-value of 1.00E-02 each in *V. destructor*'s and *V. jacobson'*s genome.

Another new neuropeptide recently described as a spider venom (agatoxin) like neuropeptide was identified in *Varroa* mites. This peptide is also described in insects including *Apis mellifera, Thermobia domestica,* and *Blatella germanica* (Sturm et al., 2016). It is believed to have highly conserved homologs (Genersch et al.) in many arthropods. Despite being a toxin in spiders, tissue profile studies suggest no association of agatoxin-like peptide (ALP) with the venom gland in *A. mellifera*. Currently, there is very little information about the role of ALP in insects. It would be interesting to know how this peptide regulates the physiology of the *Varroa* mite.

Incomplete annotations:

Neuropeptide genes found in *V. jacobsoni* but not in *V. destructor*

An unexpected result from the blast search was the identification of the gene sequence for the orcokinin B and myosuppressin neuropeptides from *V. jacobsoni*'s genomic scaffold but not from *V. destructor*'s genomic scaffold as well as from TSA, and SRA database.

The orcokinin (OK) neuropeptide was first identified in crayfish based on their intense myotropic activities in the hindgut (Stangier et al., 1992). Significantly less is known about its role in insects except for a couple of reports which suggest its potential role in controlling the circadian locomotory action in cockroaches (Hofer et al., 2005), and stimulation of prothoracic glands in silkworms (Yamanaka et al., 2011). In insects, including *Anopheles gambiae* and *A. mellifera*, orcokinin neuropeptide occurs in two different forms (OKA and OKB) through alternative splicing of the open reading frame (Sterkel et al., 2012). However, in arachnids including *I. scapularis* and *Rhipicephalus microplus* (Christie, 2008), and in crustaceans including *Marsupenaeus japonicus*, *Penaeus monodon* and *Homarus americanus* (Christie et al., 2010; Christie et al., 2011; Sterkel et al., 2012), the predicted orcokinin precursors encode both OKA and OKB peptides. In the experiment, ithe sequence containing the full-length open reading frame for the orcokinin precursor that encodes only the OKB was identified in the TSA database of *V. jacobsoni,* but not from the *V. destructor* genome. No other gene sequence that encodes both OKA and OKB or OKA also in the *V. jacobsoni* genome could be identified. The gene sequence found in *V. jacobsoni* was used to search for the orcokinin gene in the *V. destructor* genome, but there was no significant match.

Myosuppressin neuropeptide is a decapeptide with a consensus sequence of X1DVX2HX3FLRFamide (X-Asp-X-X-His-X-Phe-Leu-Arg-Phe-amide)(Egerod et al., 2003) in most arthropod species. The myosuppressin was first reported from cockroaches (Holman et al., 1986) for its inhibitory activity on visceral muscles. A partial gene sequence encoding

myosuppressin was found in *V. jacobsoni's* SRA database, but no sequence was identified in *V. destructor's* genomic database. Myosuppressin receptors were found in both of the *Varroa* mite species, hence in accordance with the neuropeptide-receptor coevolution theory (Park et al., 2002), there should also be a gene sequence encoding the myosuppressin neuropeptide in *V. destructor*. The partial sequence for myosuppressin found in *V. jacobsoni* kept the conserved sequence motif of X1DVX2HX3FLRFamide. Missing sequence for the myosuppressin in *V. destructor* is likely caused by the incomplete sequence information.

Receptor gene sequence present but missing neuropeptide gene sequence

There were a number of neuropeptides that where not identified in the *Varroa* mite species, but their cognate receptors were found. These include the genes encoding insect parathyroid hormone (iPTH), prothoracicotropic hormone (PTTH), neuroparsin A, gonadulin, RYamide, and short neuropeptide F. Genes for neuropeptides like gonadulin and PTTH have low levels of conservation in their sequence among various species of Arthropods, which in turn, makes it challenging to identify for these genes through blast searches. However, GPA2, RYamide, and short Neuropeptide F genes have high levels of conservation in their sequences. As was discussed for the mysuppressin, the missing neuropeptides are likely caused by technical errors in the incomplete genome sequences of *V. destructor* and *V. jacobsoni.* Alternatively, rapid evolutions of the neuropeptide genes may have caused the unsuccessful blast search in the search algorithms.

Missing both genes for neuropeptide and receptor:

Adipokinetic hormone (AKH), allatotropin, corticotropin-releasing factor-like DH37/47 (CRF-like), glycoprotein hormone 2-alpha subunit (GPA2), ecdysis triggering hormone (ETH), and leucokinin were amongst the neuropeptides for which no neuropeptide and cognate receptor

gene was found in the *Varroa* mite genome. Interestingly, AKH is also not found in *I. scapularis.* However, previous reports suggest the presence of five AKH-ACP-like receptor genes in the genome of *T. urticae* (Veenstra et al., 2012), another arachnids. The cognate receptor of AKH was be identified in the *Varroa* mite, which supports the true-loss based on the ligand-receptor coevolution theory. Even though both neuropeptide and receptor gene of AKH were not found, the presence of several receptor genes in the genome of *T. urticae* casts doubt on whether is it a true absence of AKH signaling. Altogether, the AKH signaling system may have been lost in parasitiformes of arachnids although this hypothesis needs to be tested with further complete sequences and data from closely related other taxa.

Similarly, the neuropeptide allatotropin and its cognate receptor gene were missing in *Varroa* mite species although they were found in the other arachnid species, including *I. scapularis* and *T. urticae*. After conducting the genomic survey on several arachnids, the allatotropin receptor gene was found. Even if both the neuropeptide and the cognate receptor genes of allatotropin signaling were not found, it would still be difficult to conclude that it is a true absence in *Varroa* mites. Genes for other neuropeptides, including ETH, GPA2, leucokinin, and CRF-like, in *Varroa* mites were also found in at least one of the two reference arachnid species used in this study, *I. scapularis* and *T. urticae*. Thus, the sudden losses of these neuropeptide genes from the *Varroa* mite are unlikely true and need further confirmation.

Neuropeptide gene missing and the receptor is unknown:

Novel neuropeptide genes, including PaOGS36577, Hansolin, RFLamide were recently reported in the stick insect, *Carausius morosus*, and American cockroach, *Periplanata americanum* (Zeng et al., 2020; Liessem et al., 2018). These neuropeptide genes were not found in the various databases of the *Varroa* mite. These neuropeptide signalings were recently

discovered in insects and their cognate receptors are yet unidentified: whether those are G protein-coupled receptors, guanyl cyclase receptors, tyrosine kinase type or other novel typoes. So, at this moment, it is very difficult to draw any conclusions on the absence of these genes.

References

1. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., Walter, P., 2013. Molecular biology of the cell. 2015. Garland, New York, 139-194.
2. Ali, N., Datta, S.K., Datta, K., 2010. RNA interference in designing transgenic crops. GM crops 1, 207-213.
3. Armenteros, J.J.A., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature biotechnology 37, 420-423.
4. Bendena, W.G., 2010. Neuropeptide physiology in insects, Neuropeptide Systems as Targets for Parasite and Pest Control. Springer, pp. 166-191.
5. Burbach, J.P.H., 2011. What are neuropeptides?, Neuropeptides. Springer, pp. 1-36.
6. Caers, J., Boonen, K., Van Den Abbeele, J., Van Rompay, L., Schoofs, L., Van Hiel, M.B., 2015. Peptidomics of neuropeptidergic tissues of the tsetse fly Glossina morsitans morsitans. Journal of The American Society for Mass Spectrometry 26, 2024-2038.
7. Christie, A.E., 2008. Neuropeptide discovery in Ixodoidea: an in silico investigation using publicly accessible expressed sequence tags. General and comparative endocrinology 157, 174-185.
8. Christie, A.E., Durkin, C.S., Hartline, N., Ohno, P., Lenz, P.H., 2010. Bioinformatic analyses of the publicly accessible crustacean expressed sequence tags (ESTs) reveal numerous novel neuropeptide-encoding precursor proteins, including ones from members of several little studied taxa. General and comparative endocrinology 167, 164-178.
9. Crooks, G.E., Hon, G., Chandonia, J.-M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome research 14, 1188-1190.
10. Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32, 1792-1797.
11. Egerod, K., Reynisson, E., Hauser, F., Cazzamali, G., Williamson, M., Grimmelikhuijzen, C.J., 2003. Molecular cloning and functional expression of the first two specific insect myosuppressin receptors. Proceedings of the National Academy of Sciences 100, 9808-9813.
12. Genersch, E., Von Der Ohe, W., Kaatz, H., Schroeder, A., Otten, C., Büchler, R., Berg, S., Ritter, W., Mühlen, W., Gisder, S., 2010. The German bee monitoring project: a long term study to understand periodically high winter losses of honey bee colonies. Apidologie 41, 332-352.
13. Hauser, F., Cazzamali, G., Williamson, M., Blenau, W., Grimmelikhuijzen, C.J., 2006. A review of neurohormone GPCRs present in the fruitfly Drosophila melanogaster and the honey bee Apis mellifera. Progress in neurobiology 80, 1-19.
14. Hewes, R.S., Taghert, P.H., 2001. Neuropeptides and neuropeptide receptors in the Drosophila melanogaster genome. Genome research 11, 1126-1142.

15. Holman, G., Cook, B., Nachman, R., 1986. Primary structure and synthesis of a blocked myotropic neuropeptide isolated from the cockroach, Leucophaea maderae. Comparative biochemistry and physiology. C, Comparative pharmacology and toxicology 85, 219-224.
16. Hofer, S., Dircksen, H., Tollbäck, P., Homberg, U., 2005. Novel insect orcokinins: characterization and neuronal distribution in the brains of selected dicondylian insects. Journal of Comparative Neurology 490, 57-71.
17. Kapp, K., Schrempf, S., Lemberg, M.K., Dobberstein, B., 2009. Post-targeting functions of signal peptides. Protein Transport into the Endoplasmic Reticulum, 1-16.
18. Li, C., Kim, K., 2008. Neuropeptides. WormBook: the online review of C. elegans biology, 1.
19. Liessem, S., Ragionieri, L., Neupert, S., Büschges, A., Predel, R., 2018. Transcriptomic and neuropeptidomic analysis of the stick insect, Carausius morosus. Journal of proteome research 17, 2192-2204.
20. Nässel, D.R., 2018. Substrates for neuronal cotransmission with neuropeptides and small molec Nässel, D.R., Zandawala, M., 2019. Recent advances in neuropeptide signaling in Drosophila, from genes to physiology and behavior. Progress in neurobiology 179, 101607.ule neurotransmitters in Drosophila. Frontiers in cellular neuroscience 12, 83.
21. Park, Y., Kim, Y.-J., Adams, M.E., 2002. Identification of G protein-coupled receptors for Drosophila PRXamide peptides, CCAP, corazonin, and AKH supports a theory of ligand-receptor coevolution. Proceedings of the National Academy of Sciences 99, 11423-11428.
22. Price, D.A., Greenberg, M.J., 1977. Structure of a molluscan cardioexcitatory neuropeptide. Science 197, 670-671.
23. Stangier, J., Hilbich, C., Burdzik, S., Keller, R., 1992. Orcokinin: a novel myotropic peptide from the nervous system of the crayfish, Orconectes limosus. Peptides 13, 859-864.
24. Sterkel, M., Oliveira, P.L., Urlaub, H., Hernandez-Martinez, S., Rivera-Pomar, R., Ons, S., 2012. OKB, a novel family of brain-gut neuropeptides from insects. Insect biochemistry and molecular biology 42, 466-473.
25. Sturm, S., Ramesh, D., Brockmann, A., Neupert, S., Predel, R., 2016. Agatoxin-like peptides in the neuroendocrine system of the honey bee and other insects. Journal of Proteomics 132, 77-84.
26. Taghert, P.H., Schneider, L.E., 1990. Interspecific comparison of a Drosophila gene encoding FMRFamide-related neuropeptides. Journal of Neuroscience 10, 1929-1942.
27. Veenstra, J.A., 2000. Mono-and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America 43, 49-63.
28. Yamanaka, N., Roller, L., Žitňan, D., Satake, H., Mizoguchi, A., Kataoka, H., Tanaka, Y., 2011. Bombyx orcokinins are brain-gut peptides involved in the neuronal regulation of ecdysteroidogenesis. Journal of Comparative Neurology 519, 238-246.
29. Zeng, H., Qin, Y., Du, E., Wei, Q., Li, Y., Huang, D., Wang, G., Veenstra, J.A., Li, S., Li, N., 2020. Genomics-and Peptidomics-Based Discovery of Conserved and Novel Neuropeptides in the American Cockroach. Journal of Proteome Research.

Figures:

(2A)

Varroa destructor

MQFRINAFLPLGPATLCVLLALTFLSAVPTAAQDDRDRSGVAEMVQSLSPEEQLFYDVLVKRPNANHRYGFGLGKRS
PEPTPAQPFQDEPTFSTFKRRQYNFGLGKRPWPMSDNEEYRKRKYNFGLGKRSE

Varroa jacobsoni

MQFRINAFLPLGPATLCVLLALTFLSAVPTAAQDDRDRSGVAEMVQSLSPEEQLFYDVLVKRPNANHRYGFGLGKRS
PEPTPAQPFQDEPTFSTFKRRQYNFGLGKRPWPMSDNEEYRKRKYNFGLGKRSE

(2B)

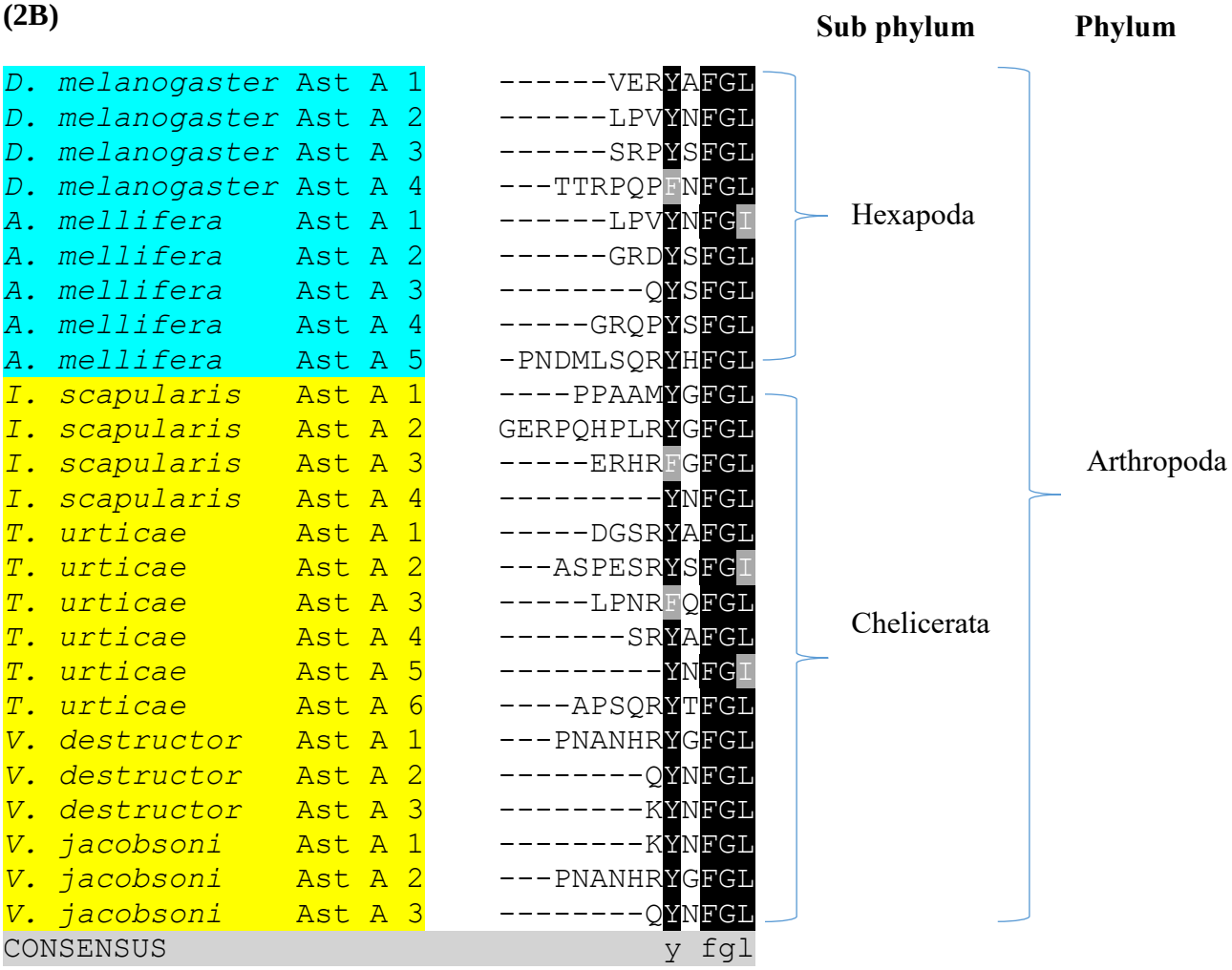


Figure 2.2 (A) The putative preprohormone sequence of Allatostatin A gene for *Varroa destructor* and *Varroa jacobsoni*. The sequence marked in the green color indicate the signal peptide necessary for initiation of the secretory pathway, mature peptide is indicated in the yellow color, the amidation site is marked with pink color and the amino acids in the red color represents the monobasic or dibasic cleavage sites. (B) Multiple sequence alignment of multiple mature peptides of Ast A gene

(3A)

Drosophila melanogaster

MAHTKTRRTYGFLMVLLILGSACGNLVASGSAGSPPSNEPGGGGLSEQVVLDQLSESDLYGNNKRAWQSLQSSWGKR
SSSGDVSDPDIYMTGHFVPLVITDGTNTIDWDTFERLASGQSAQQQQQQPLQQQSQSGEDFDDLAGEPDVEKRAWKS
MNVAWGKRRQAQGWNKFRGAWGKREPTWNNLKGMWGKRDQWQKLHGGWGKRSQLPSN

Nephrops norvegicus

MMTVQQMCRPWALLVVVLVAGATQVSSSSSSSPQQDDPASSPSHIEEKRVGWSSMRGTWGKRPHLEDAQLDAAEVKR
TNWNKFHGSWGKRGEELQAAEDKRTNWNKFQGSWGKRADDLADAELQAAEDKRTNWNKFQGSWGKRADDLADAELQA
AEDKRTNWNKFQGSWGKRADDLADA

Daphnia pulex

MQFWQCPLLLMVSLIAAINTQQTPSQRLEPNQVAGLVELHQQLEQPREHQQQQQHHQQQPEQPAQADNKDYAPSPAV
LLQLTPSDWKNTQDKRNNWNRMQGMWGKRSQQQSDESALSEMTPPRMVKRAWSDLSQQGWGKRSWTQLHGVWGKRRW
DQLHGAWGKRTPDQLEDDSKAEQPENSQEDEDQQSSSEVEREEANDSDDTVENSKRSGWNKMQGVWGKRSSSSSKTN
SGPAIEGMGNNDLLLLISGTGDQLYQQREDDQAKANVEDAGNEESVDKRGWNQLQGVWGKRALSAMAAGYKRNWNNL
RGAWGKREIPAAIAKGMEWSRKRESGWNNLKGLWG

Varroa destructor

MSTALLTVALVIAVCAVGTFGKLDAESPPSAPSPVEYPPQYFDAPLEAEYVLLKKADVPPAPWNRLYNDWGKRADNW
KNLNHLWGKRSATLPTRWDKRPQPQWNELSGYWGKRSAQ

Varroa jacobsoni

MSTALLTVALVIAVCAVGTFGKLDAESPPSAPSPVEYPPQYFDAPLEAEYVLLKKADVPPAPWNRLYNDWGKRADNW
KNLNHLWGKRSATLPTRWDKRPQPQWNELSGYWGKRSAQ

Limulus polyphemus

MRCSTFKTSCILIVLYFTALVKSKSNEVNSALEEALTNARNEDDQKRDWNNLSGMWGKRGWNNLSGMWGKRGSKWNN
LSGMWGKRGSSWNNLSGMWGKRGSNWNNLSGMWGKRGSDWNNLSGMWDKRSPSWNDLSGMWGKRGWNNLSGMWGKRN
TNWNNLKGLWGKRGDSSFNEGGPSWRITDAGI

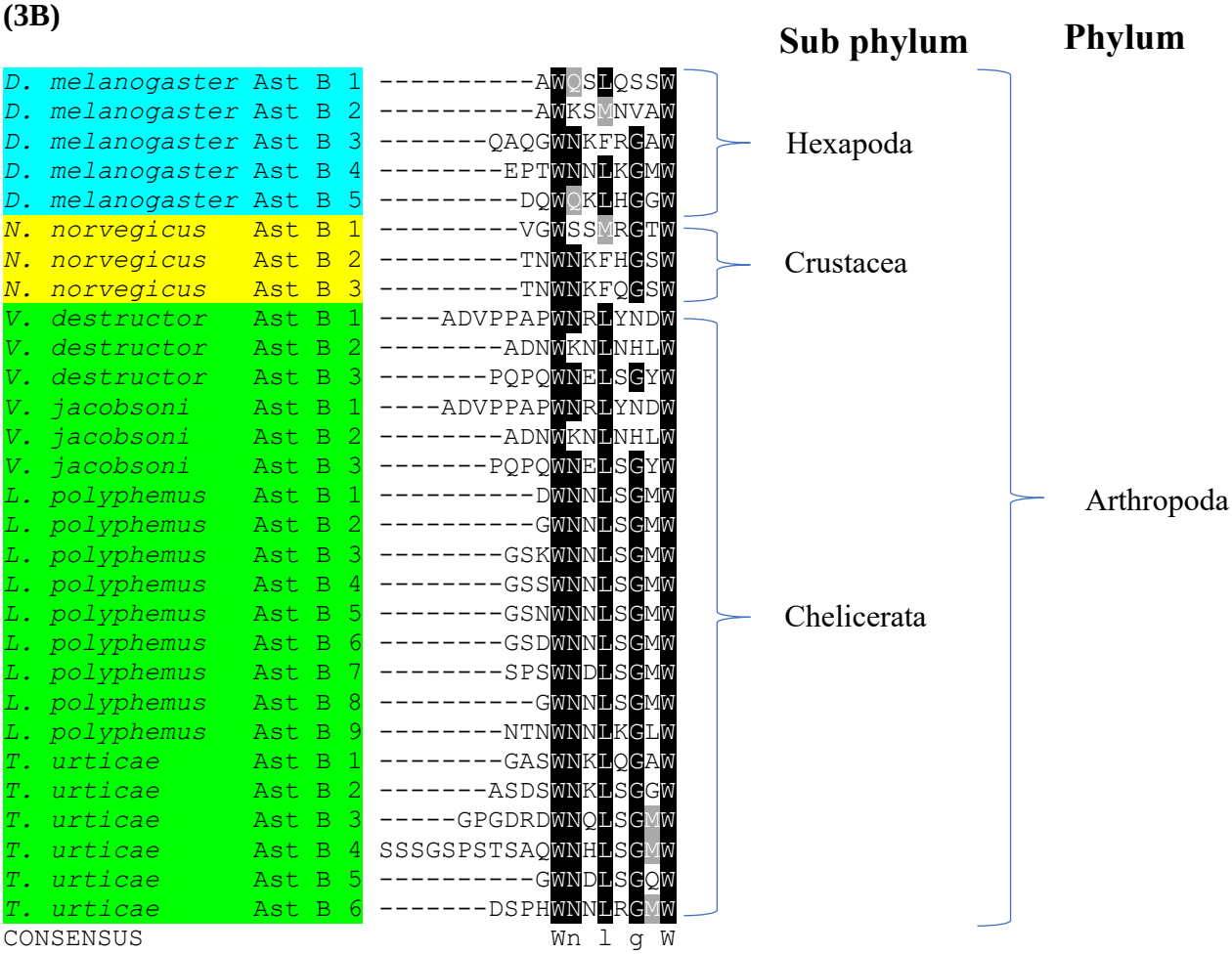
Galendromus occidentalis

MFMKTCVVSAAFMALIASTMAKPDPESANVASAPQEYFDAPVDAEYIILKKADAAPAHWNRLYNDWGKRADQWKNLN
HMWGKRSAPAPNRWDKRPQPQWNELSGYWGKRSSA

Tetranychus urticae

MVKSSEVCCYFQINLLLLLTIWSVSLISSVSLAPESVTNLDQYATDPDSIGTASHLSLPTSSSSSSLLGQQHKGNKI
TVYYPNMESQDFFVDDLEDNQIMEKKGASWNKLQGAWGKRASDSWNKLSGGWGKRGPGDRDWNQLSGMWGKRSSSGS
PSTSAQWNHLSGMWGKRGWNDLSGQWGKRDSPHWNNLRGMWG

(3B)



(3C)

Varroa destructor:

```
GACTACTGAAGCGTCGTGATGTGCCGCACCGCCAATAACATCAACACGAACTGCAGCGGAGCGATGAGTACCGCGC
                                                                M  S  T  A
TGTTGACGGTTGCCCTAGTCATTGCAGTATGCGCGGTAGGTACTTTCGGAAAGTTAGACGCGGAATCACCGCCCAGC
L  L  T  V  A  L  V  I  A  V  C  A  V  G  T  F  G  K  L  D  A  E  S  P  P  S
GCACCATCTCCAGTTGAGTACCCTCCCCAATACTTCGATGCGCCCCTTGAAGCAGAGTATGTTCTTCTCAAAAAAGC
 A  P  S  P  V  E  Y  P  P  Q  Y  F  D  A  P  L  E  A  E  Y  V  L  L  K  K  A
TGACGTACCTCCAGCGCCTTGGAACCGCTTGTACAATGATTGGGGTAAAAGGGCTGATAACTGGAAGAATCTAAATC
  D  V  P  P  A  P  W  N  R  L  Y  N  D  W  G  K  R  A  D  N  W  K  N  L  N
ACCTGTGGGGCAAACGGTCAGCTACACTTCCGACCCGGTGGGACAAACGCCCTCAGCCGCAGTGGAACGAGCTATCC
H  L  W  G  K  R  S  A  T  L  P  T  R  W  D  K  R  P  Q  P  Q  W  N  E  L  S
GGTTATTGGGGAAAGCGTTCGGCCCAGTAAATGTACCGCCTGAATAGAAGTAAACAGAAATGTA
 G  Y  W  G  K  R  S  A  Q  *
CTGCAGAAGAAGGCTAGCTCTCGTTCTCACCGCTATCCGTAAGTGAAAAGGACTTCGAAGACCAGGTGTCACCATGGGGCAGTCGATTTTAGCAGTA
ACCGGATTCGTAAAGGAGCCGGCTGGAAACGATGGAAAGCATACAAAAACATACATTTTTTTGGAGGCCTAAGGCCTCTGATGGCAAAAATTGTTTA
CCGCACACATAATCAATAAAAAATACTAAAGGATATATCCAGATTCGATGACAACATCATACTGGTCATAATGTCTAATGATTATAGTCTAGCAAAG
GAATATAACTCAACTTTATCAAAAGAACGTGACTATGAACCTATTGAACAGGAATACCATTGCGAAGACGCTAATATCTTAGCGATAAATAATAAAT
ATAAGCAGCTTATTTTCTAGATTACAACATATAACGTACGCATAAGTCATTAAAATAATGAAACAAATATGGAAAGGCTAATGTCGAAACGAATGGA
AACATAGTAGGTTGAAACAAGCAATGACAGCACCAAAAACATCAAGTACTATTTAACAATCAATTATATCATTTACGTAAGAAGATACAAATAGGGA
CGGCACGGGTATTGCAAAAGGCAAAAGGGTGCTTCTCAAAGCTCTCCGATCAAGTTAGTCCTCTTTCCAGGACCGGATCGGCGCTCTGCACGAAGGT
GCAGGCCGTTTTCTAATTTCTTTGCCACGCCGTGTTGATTATCATTAGCAAGAACTGAACAACCGTCACAGCCATTAGTTAACTAGAACATATACAC
```

AAATATCACCTTCACATGAGAAGACCAAACGTTATTTTTAAAACGTTTGTTCAATTCATAGCACGATCGAGCCTATGCATGACGAAACCTCAGGGAA
AACAAAGTTCTCGTATCAACAGATCGGACAGACGAGGTAATGGATAGTGACAAGTAAGCGGATAATCGGTCATCGCCTCTTCGAATGCGACCGTAAT
CTCCGTGTGTATTGGCTTAACTGATCTCAATCATTAAGACAATCTATGACAAAATCTCATGATTGAAGACGCGTTACTGTATGGTGACATGCCGCAT
TGCACGGAGTACTAATTTTCTAGGTCTCAAGATTTTCAGCTCACTTGTTAATTCACTACTATGAGACCATATTTTTATCTACATATTGTTGTTTTGT
ATCATACAGATGTCCAAATTCAAGACGTTAACGCTCACGCTTTATGTAGATGAGGAAATTAGATTTACACTATAGGCAAGAGCTAAGTCCGCTAGCT
TACCTACTGAAATATGAAGCATGTTTATATAGCGACGATTCTGGTATTTTTTTTTGCATGTAGAGCTACAAACGACAGTGTCTAGGATTCATGCTGC
AAAAATAAACATTTTTTGATAGATCTCAATTCTCCTAAGCTTCGTTGTTACCACAGATGGAGTGACAACATCCACGAGGCCCTGGCACTGCGATTTA
TTCGACTTTTTCGGGACGTATGTTCCAAACGTTTTCGGTTCGTTAAGTATTCACTCGCGACACGGCCGTATACCAGCTTTCGTGCTCCACACTCTGC
TCGTAAAGTAAATAACACTACTCTATCGCTTATAACAAAAACCACCCCCCACTGAGCCACGGAAATGCTAATTCGAAGCTGGAAAGAGAGAACAACT
AACGATCAATAACATGTGACTGATAAAGCTTCAAAATAGTGCATAAGCTATATACGAGATAGTAACCAGGCAGGACGCAAATGATACCACACCGTAA
CCACTCTCATAGAGTCATTGGAAACACGGCTCTTATTCTACTGACGGTATTCTAACTTGCGACTTGTTCTACGCTACTTAATGCAATGAACGAAATT
GTAATTCATTTTTTAATTTTTACATATATATTTCTTTTTATCACAAATGAATAATAAAACACATTTTTCAACAATAAATAACAATTCTGAATTAAGG
AAAAATGAAAATACCAATAACTAAGGACTGCTGCGCGGAAGATACCCTATGGTAAGTTCACGTCTTCGTCTGACGAGAACAAAAGTAATTGCTAATC
CCACAGATCTTCACTAAGAGACCCTTGATTCTCCCTTCCAGCAATATTGGACCCGAACGCAGCAACTAATATCAATGACAGGGCGGACAATAATATC
ACCAAAAGCAAGGAGAGAGTCTCAATAGGGAAGAGACGACCCATAGATTGTCAAATATAACAACTACGTACTTTAGAGACACGGCGATAACAATGCA
ATGCCATAATGTTGTAAAAGTAAACATGCAAATGAAGAGAAAAAAGACAAGTAAAAA

Varroa jacobsoni:

CCGCCAACACATAAAAGGAGTTGCGCCCCTTTATTTGAGCCGAGAAAGCATAGAGTTTTGAACTAGCAAGGAGGTGC

CACAGCTTTCAAGTTTGAAAAAGATCACGCTAATCCTTGACGCAGCGCTGTCGACGAACTCCGTAACGACTACTGAA
GCGTCGTGATGTGCCGCACCGCCAATAACATCAACACGAACTGCAGCGGAGCGATGAGTACCGCGCTGTTGACGGTT
M S T A L L T V
GCCCTAGTCATTGCAGTATGCGCGGTAGGTACTTTCGGAAAGTTAGACGCGGAATCACCGCCCAGCGCACCATCTCC
A L V I A V C A V G T F G K L D A E S P P S A P S P

AGTTGAGTACCCTCCCCAATACTTCGATGCGCCCCTTGAAGCAGAGTATGTTCTTCTCAAAAAAGCTGACGTACCTC
V E Y P P Q Y F D A P L E A E Y V L L K K A D V P

CAGCGCCTTGGAACCGCTTGTACAATGATTGGGGTAAAAGGGCTGATAACTGGAAGAATCTAAATCACCTGTGGGGC
P A P W N R L Y N D W G K R A D N W K N L N H L W G
AAACGGTCAGCTACACTTCCGACCCGGTGGGACAAACGCCCTCAGCCGCAGTGGAACGAGCTATCCGGTTATTGGGG
K R S A T L P T R W D K R P Q P Q W N E L S G Y W G
AAAGCGTTCGGCCCAG**TAA**
K R S A Q *****

ATGTACCGCCTGAATAGAAGTAAACAGAAATGTACTGCAGAAGAAGGCTAGCTCTCGTTCTCACCGCTATCCGGAAGTGAAAAGGACTTCGAAGACC
AGGTGTCACCATGGGGCAGTCGATTTTAGCAGTAACCGGATTCGTAAAGGAGCCGGCTGGAAACGATGGAAAGCATACAAAAACATACATTTTTTTG
GAGGCCTAAGGCCTCTGATGGCAAAAATTGTTTACCGCACACATAATCAATAAAAAATACTAAAGGATATATCCAGATTCGATGACAACATCATACT
GGTCATAATGTCTAATGATTATAGTCTAGCAAAGGAATATAACTCAACTTTATCAAAAGAACGTGACTATGAACCTATTGAACAGGAATACCATTGC
GAAGACGCTAATATCTTAGCGATAAATAATAAATATAAGCAGCTTATTTTCTAGATTACAACATATAACGTACGCATAAGTCATTAAAATAATGAAA
CAAATATGGAAAGGCTAATGTCGAAACGAATGGAAATATAGTAGGTTGAAACAAGCAATGACAGCACCAAAAACATCAAGTACTATTTAACAATCAA
TTATATCATTTACGTAAGAAGATACAAATAGGGACGGCACGGGTATTGCAAAAGGCAAAAGGGTGCTTCTCAAAGCTCTCCGATCAAGTTAGTCCTC
TTTCCAGGACCGGATCGGCGCTCTGCACGAAGGTGCAGGCCGTTTTCTAATTTCTTTGCCACGCCATGTTGATTATCATTAGCAAGAACTGAACAAC
CGTCACAGCCATTAGTTAACTAGAACATATACACAAATATCACCTTCACATGAGAAGACCAAACGTTATTTTTAAAACGTTTGTTCAATTCATAGCA
CGATCGAGCCTATGCATGACGAAACCTCAGGGAAAACAAAGTTCTCGTATCAACAGATCGGACAGACGAGGTAATGGATAGTGACAAGTAAGCGGAT
AATCGGTCATCGCCTCTTCGAATGCGACCGTAATCTCCGTGTGTATTAGCTTAACTGATCTCAATCATTAAGACAATCTATGACAAAATCTCATGAT
TGAAGACGCGTTACTGTATGGTGACATGCCGCATTGCACGGAGTACTAATTTTCTAGGTCTCAAGATTTTCAGCTCACTTGTTAATTCACTACTATG
AGACCATATTTTTATCTACATATTGTTGTTTTGTATCATACAGATGTCCAAATTCAAGACGTTAACGCTCACGCTTTATGTAGATGAGGAAATTAGA
TTTACACTATAGGCAAGAGCTAAGTCCGCTAGCTTACCTACTGAAATATGAAGCATGTTTATATAGCGACGATTCTGGTATTTTTTTTTTTGCATGT
AGAGCTACAAACGACAGTGTCTAGGATTCATGCTGCAAAAATAAACATTTTTTGATAGATCTCAATTCTCCTAAGCTTCGTTGTTACCACAGATGGA
GTGACAACATCCACGAGGCCCTGGCACTGCGATTTATTCGACTTTTTCGGGACGTATGTTCCAAACGTTTTCGGTTCGTTAAGTATTCACTCGCGAC
ACGGCCGTATACCAGCTTTCGTGCTCCACACTCTGCTCGTAAAGTAAATAACACTACTCTATCGCTTATAACAAAAACCACCCCCCACTGAGCCACG
GAAATGCTAATTCGAAGCTGGAAAGAGAGAACAACTAACGATCAATAACATGTGACTGATAAAGCTTCAAAATAGTGCATAAGCTATATACGAGATA
GTAACCAGGCAGGACGCAAATGATACCACACCGTAACCACTCTCATAGAGTCATTGGAAACACGGCTCTTATTCTACTGACGGTATTCTAACTTGCG
ACTTGTTCTACGCTACTTAATGCAATGAACGAAATTGTAATTCATTTTTTAATTTTTACATATATATTTCTTTTTATCACAAATGAATAATAAAACA
CATTTTTCAACAATAAATAACAATTCTGAATTAAGGAAAAATGAAAATACCAATAACTAAGGACTGCTGCGTGGAAGATACCCTATGGTAAGTTCAC
GTCTTCGTCTGACGAGAACAAAAGTAATTGCTAATCCCACAGATCTTCACTAAGAGACCCTTGATTCTCCCTTCCAGCAATATCGGACCCGAACGCA
GCAACTAATATCAATGACAGGGCGGACAATAATATCACCAAAAGCAAGGAGAGAGTCTCAATAGGGAAGAGACGACCCATAGATTGTCAAATATAAC
AACTACGTACTTTAGAGACACGGCGATAACAATGCAATGCCATAATGTTGTAAAAGTAAACATGCAAATGAAGAGAAAAAAGACAAGTAAAAATCAG
AACGTCGACACACAAACGAATGACAGACACATACTGAGTTATAAATGGAAGCTAATAAACAATCAGATTAAAAGCAAGAAAAAACAAGCGGCACGAG
TGTGAAACCCAAAAGAATTCCGTAAAGATATTTGTTTACATAACGAAGCAGTCAAGCAAGAAATATAGCAGTCGATCAATGACGGGGTACTTGCTAT
CAAGCCAAAAATGTTGTTATTATATGAAATCTATATATAGTAGATAAAAACGTATAACATAAGCACACGATTATTATGCTAACGTACTTAAGATATG
ATAATGATAATAATATGTGTTATACATATATATGATGATAAAAAGATAGATATAATGGTACGAGAAAGTATTTGTAGATTATTATTATCAAACAAAA
AGAAAGGGTGTGGTAACATCGAGGGAACGAAAACTGCGAGCCAAAGCACGTTGCAAGATTAAAGGCACACCTTCGAACCTCCTAAAGAAAGGAAGTA
ACGGAAAAAATGAAAAAAAAAAGAAACAGGCCACAAGAAAGAGAACAGAGTACTGCTGATAAAATTGCTGATGTCTTTTTGTTCGGCCAATCTCCAA

GTACTTTGAAGAAGTGCACGATTTATGACTCCGTTGATTAAGAAGTGCACTTTTATTGAAAAAAAAGACCTTAGAAATATATGGAACTAATCGTTTC
TACTAGGACTATAATATCCCGCAGCTTATAATAATATAATCCAGCTAAAAGCTCATCTTAGGTACTGGGATTAGGAAAAACATTATTCTACGTCTGA
TGGAAAGTGTATTGATACTGGCACTAATCAAGCGTAGATTCAGAGGGGTTTTCGGTTCGTTAAGTATTCACTCGCGACACGGCCGTATACCAGCTTT
CGTGCTCCACACTCTGCTCGTAAAGTAAATAACACTACTCTATCGCTTATAACAAAAACCACCCCCCACTGAGCCACGGAAATGCTAATTCGAAGCT
GGAAAGAGAGAACAACTAACGATCAATAACATGTGACTGATAAAGCTTCAAAATAGTGCATAAGCTATATACGAGATAGTAACCAGGCAGGACGCAA
ATGATACCACACCGTAACCACTCTCATAGAGTCATTGGAAACACGGCTCTTATTCTACTGACGGTATTCTAACTTGCGACTTGTTCTACGCTACTTA
ATGCAATGAACGAAATTGTAATTCATTTTTTAATTTTTACATATATATTTCTTTTTATCACAAATGAATAATAAAACACATTTTTCAACAATAAATA
ACAATTCTGAATTAAGGAAAAATGAAAATACCAATAACTAAGGACTGCTGCGTGGAAGATACCCTATGGTAAGTTCACGTCTTCGTCTGACGAGAAC
AAAAGTAATTGCTAATCCCACAGATCTTCACTAAGAGACCCTTGATTCTCCCTTCCAGCAATATCGGACCCGAACGCAGCAACTAATATCAATGACA
GGGCGGACAATAATATCACCAAAAGCAAGGAGAGAGTCTCAATAGGGAAGAGACGACCCATAGATTGTCAAATATAACAACTACGTACTTTAGAGAC
ACGGCGATAACAATGCAATGCCATAATGTTGTAAAAGTAAACATGCAAATGAAGAGAAAAAAGACAAGTAAAAATCAGAACGTCGACACACAAACGA
ATGACAGACACATACTGAGTTATAAATGGAAGCTAATAAACAATCAGATTAAAAGCAAGAAAAAACAAGCGGCACGAGTGTGAAACCCAAAAGAATT
CCGTAAAGATATTTGTTTACATAACGAAGCAGTCAAGCAAGAAATATAGCAGTCGATCAATGACGGGGTACTTGCTATCAAGCCAAAAATGTTGTTA
TTATATGAAATCTATATATAGTAGATAAAAACGTATAACATAAGCACACGATTATTATGCTAACGTACTTAAGATATGATAATGATAATAATATGTG
TTATACATATATATGATGATAAAAAGATAGATATAATGGTACGAGAAAGTATTTGTAGATTATTATTATCAAACAAAAAGAAAGGGTGTGGTAACAT
CGAGGGAACGAAAACTGCGAGCCAAAGCACGTTGCAAGATTAAAGGCACACCTTCGAACCTCCTAAAGAAAGGAAGTAACGGAAAAAATGAAAAAAA
AAAGAAACAGGCCACAAGAAAGAGAACAGAGTACTGCTGATAAAATTGCTGATGTCTTTTTGTTCGGCCAATCTCCAAGTACTTTGAAGAAGTGCAC
GATTTATGACTCCGTTGATTAAGAAGTGCACTTTTATTGAAAAAAAAGACCTTAGAAATATATGGAACTAATCGTTTCTACTAGGACTATAATATCC
CGCAGCTTATAATAATATAATCCAGCTAAAAGCTCATCTTAGGTACTGGGATTAGGAAAAACATTATTCTACGTCTGATGGAAAGTGTATTGATACT
GGCACTAATCAAGCGTAGATTCAGAGGGGACTGATATAAATGTTCCAACTAATAATACCTGATAATGATTTAATATTGACGCAAAGGTCTGGTGTAA
AAAAGTTACCAAAATGACGAATTAATTCGTTTAATTTTGTTTTGTAGTCAATATCTCAGTCAGTGCGATCGGATTGCACCTTACTTACTTAAATCTT
ATCTAGCTTTGTTTATCTGTCAGTATAATGACGGAATTTATTTTTTATCAGACAAAAATTGCACAAACGTTAAAATAATAATTAATTATGCCATTCT
CAAAGAAATTACCACAATCATCGCTAAATGAATCGACTAATGCTCATTATATAGCCTTGCAATTCTTCCTTCCATAGATTTTAAGAAAGGCCTTTAC
TAACCTATGTGTGTTTATATATTATTGCAACCAAATTATGCTGCGTTTGAAAATACAAAGCCTA

Figure 2.3 (A) Preprohormone sequence of genes coding for Allatostatin B from various species in Phylum Arthropoda, indicating the multiple mature peptide in a gene sequence. The color scheme is same as indicated above: green color = Signal peptide, yellow color = mature peptides, red color = monobasic and dibasic cleavage sites. The sequence with red font represents accessory proteins. (B) Multiple sequence alignment mature peptide repeats, representing the conservation of the amino acid at particular position in respect with C terminal. The color distinction represents the Class distinction of species to which they belong. The consensus sequence is given at the end of the alignment. (C) Figure represents the gene structure for both *V. destructor* and *V. jacobsoni*. The signal peptide is indicated in bold and italics font. Red color represents the dibasic cleavage sites, yellow color of sequence represents the mature peptide, pink color represents amidation site.

(4A)

Ixodes scapularis
MNSWKAFFMFGTLLVMAVMMNMACAAYRKPPFNGSIFGKRSRADLNNADVKYAMCEAVWDTCTQWFPITQ
DGAQ

Araneus ventricosus
MSKLLKAIFVLVLCTLVVILPAASRKPPFNGSIFGKRFVRDMDVDNRGGSDFRHPACEYVYDACNQWYMN
GQDSQ

Varroa destructor
MVRCLALYLMCASILHFTDVNYRKPPFNGSIFGKRNNGLVAGRVAGQDDRGVCEVLLDACGQLLQSVIDQ
QGV

Varroa jacobsoni
MVRCLALYLMCASILHFTDVNYRKPPFNGSIFGKRNNGLVAGRVAGQDDRGVCEVLLDACGQLLQSVIDQ
QGV

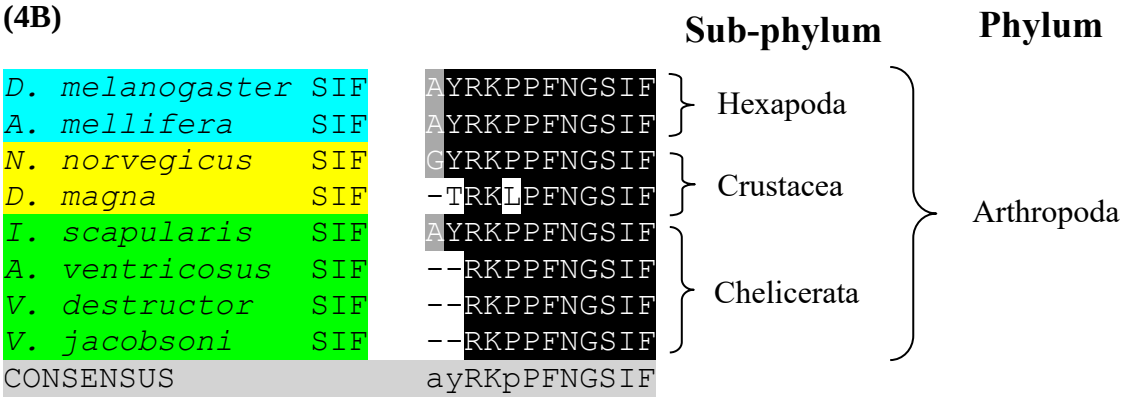
Nephrops norvegicus
MSVQMRVVVALAVVLVIVAVLTDPVSAGYRKPPFNGSIFGKRAGADPLFEPGKGLASVCQVAVEACAAWF
PVPEKK

Daphnia magna
MMRPSFMIVMVCFVFIFSFWGQLVEGTRKLPFNGSIFGKRSNPGAGKHAASPNFQLLCDAAMNACSDWLP
IGSN

Drosophila melanogaster
MALRFTLTLLLVTILVAAILLGSSEAAYRKPPFNGSIFGKRNSLDYDSAKMSAVCEVAMEACPMWFPQND
SK

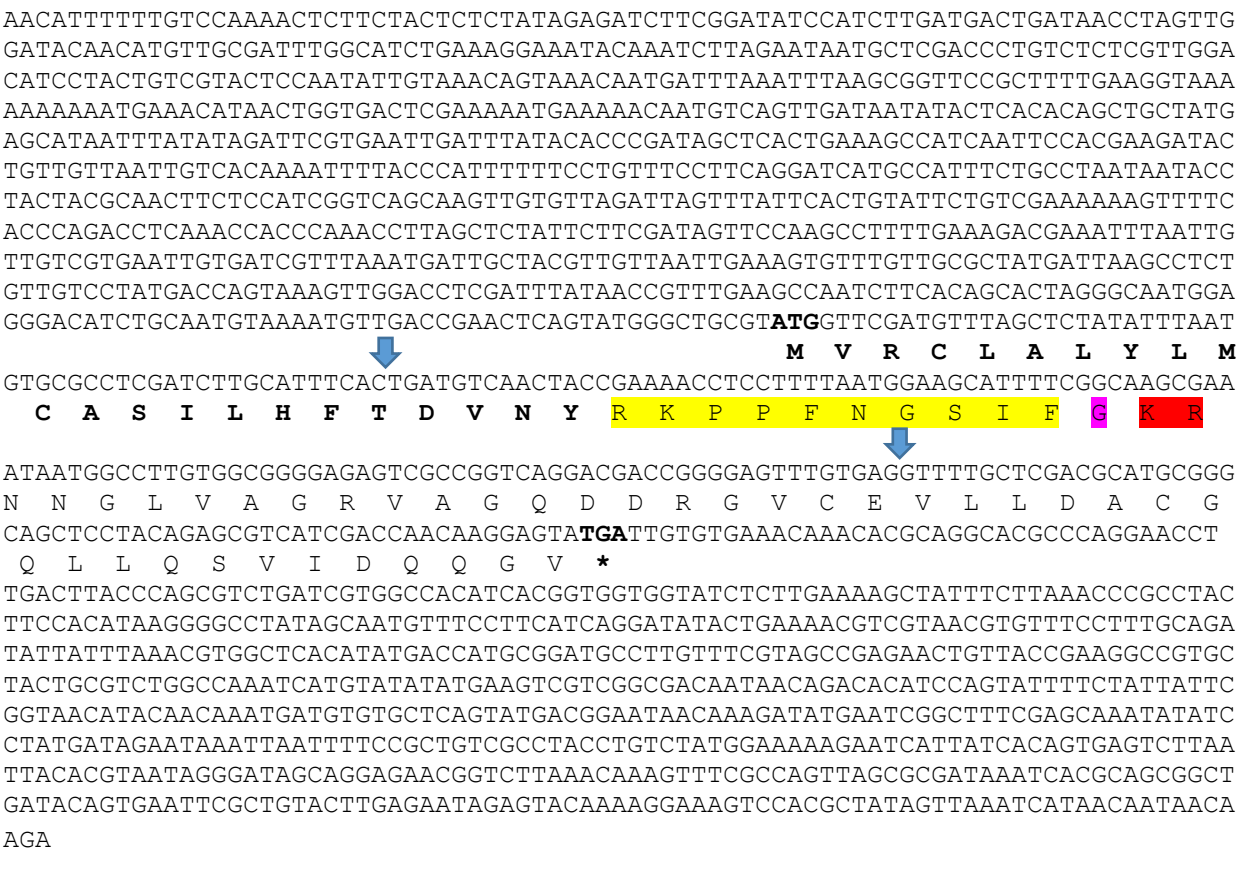
Apis mellifera
MVSTRLVLAVVAALFVLAISVDAAYRKPPFNGSIFGKRSNTITDYEITSRAMSSVCEVVSETCNAWLSRQ
DSN

(4B)



(4C)

Varroa destructor



Varroa jacobsoni

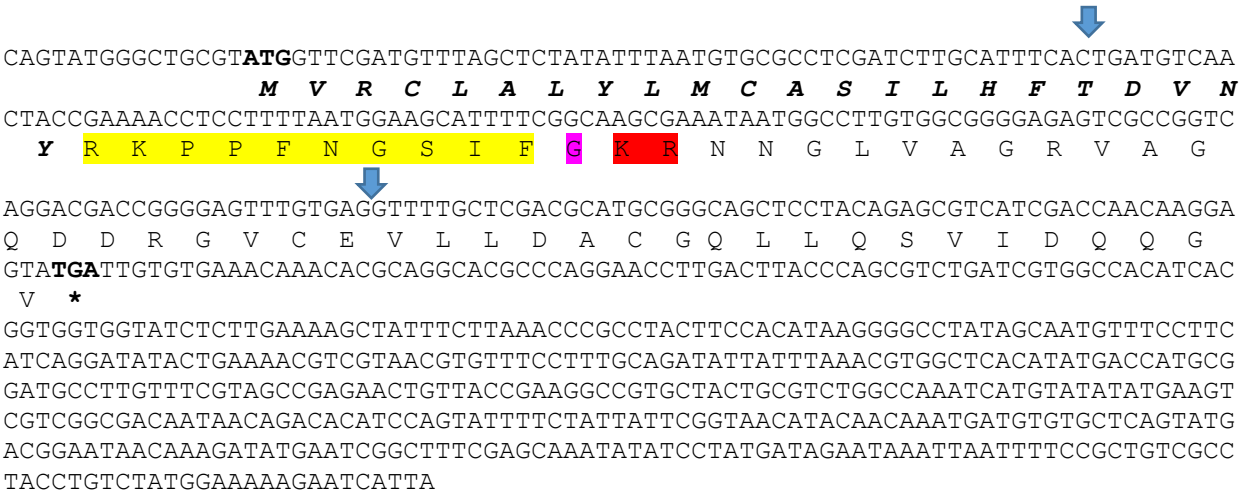


Figure 2.4 (A) Preprohormone gene sequence encoding SIFamide neuropeptide among various species in the Phylum Arthropoda. The color scheme is same as indicated above: green color = Signal peptide, yellow color = mature peptides, red color = monobasic and dibasic cleavage sites. The sequence with red font represents accessory proteins. (B) Multiple sequence alignment mature peptide repeats, representing the conservation of the amino acid at particular position in respect with C terminal. The color distinction

represents the Class distinction of species to which they belong, blue is for Class for Hexapoda, yellow is for Class for Crustacea and green is for Class Chelicerata. The consensus sequence is given at the end of the alignment. (C) Figure represents the gene structure for both *V. destructor* and *V. jacobsoni.* The signal peptide is indicated in bold and italics font. Red color represents the dibasic cleavage sites, yellow color of sequence represents the mature peptide, pink color represents amidation site. Arrow represents the positioning of introns

(5A)

Varroa destructor:

MKCVLPPLTGLLACGLLGGLINSTESAQAARFAASPDAARERSATTRLGDDLSLPNVRSDARYPSSSAQQSFGGSQA
SIVDTFENCAALFQFALDNLSPEERQTALANWCSKALSKNIEDVNDDSMLIGNSEMLLGKRAVKIQKFMHFGRKRGE
LDAFSTGSAPSAVDMDKHAGQNRMLHFGKRYNTFADYQDAPDRNEALRQLDEPAKRAGQNRMLHFGKRSPNRFMHFG
KRNPDSQFLHFGKKIFDQSDFESYPNKHVNRFMHFGKRDDEQADAATNLEHALDDDFVKKLDNNRFMHFGKKRAESQ
KYMHFGKRSSNELGLDSDSTASSYMDRLLKRNSGKNRYVHFGKKSTNRFMHFGKRAELGQELGSGTFPVDVEDNSFL
VDQQDQLHERRRRSILSDADGAGFRVEFGLDTKEIRPASQKRFTITYPSPIFQQALLRHRLDLLDKVDVEDSDGGLN
GVSGPLGQSGSGPLNQRQPNTATFIGSEDLLDRMGRSEQFQIPQPRVQSRWTRPDRNVFLHFG

Varroa jacobsoni:

MKCVLPPLTGLLACGLLGGLINSTESAQAARFAASPDAARERSATTRLGDDLSLPNVRSDARYPSSSAQQSFGGSQA
SIVDTFENCAALFRFALDNLSPEERQTALANWCSKALSKNIEDVNDDSMLIGNSEMLLGKRAVKIQKFMHFGRKRGE
LDAFSTGSAPSAVDMDKHAGQNRMLHFGKRYNTFADYQDAPDRNEALRQLDEPAKRAGQNRMLHFGKRSPNRFMHFG
KRNPDSQFLHFGKKIFDQSDFESYPNKHVNRFMHFGKRDDEQADAATDLEHALDDDFVKKLDNNRFMHFGKKRAESQ
NDSTASSYMDRLLKRNSGKNRYVHFGKKSTNRFMHFGKRAELGQELGSGTFPVDVEDNSFLVDQQDQLHERRRRSIL
SDADGAGFRVEFGLDTKEIRPASQKRFTITYPSPIFQQALLRHRLDLLDKVDVEDSDGGLNGVSGPLGQSGSGPLNQ
RQPNTATFIGSEDLLDRMGRSEQFQIPQPRVQSRWTRPDRNVFLHFG

(5B)

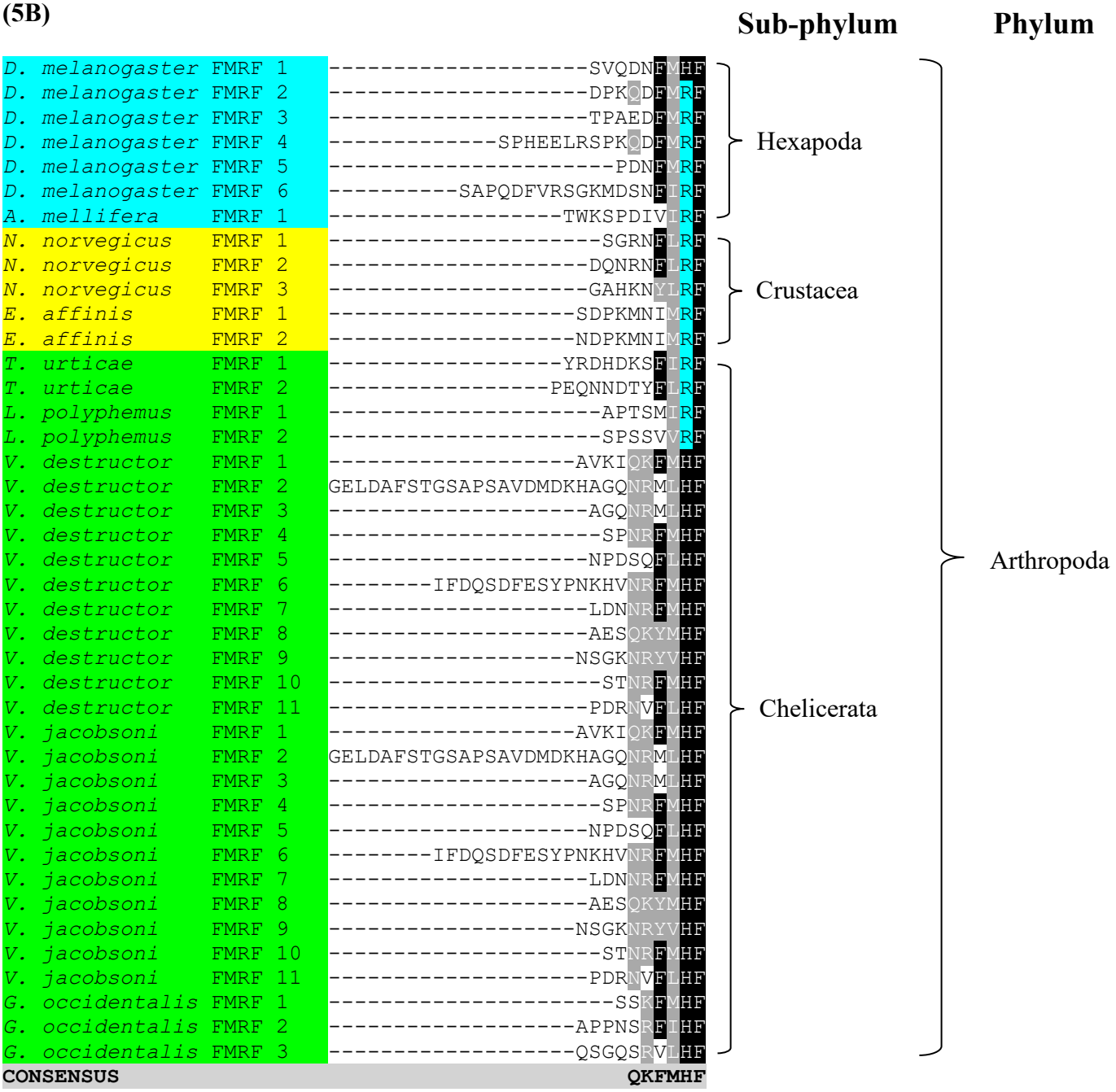


Figure 2.5 (A) Preprohormone sequence of FMRFamide gene with FMHF C terminal motif sequence. The color schemes is as follows: green color = Signal peptide, yellow color = mature peptides, red color = monobasic and dibasic cleavage sites. The sequence with red font represents accessory proteins.(B) Multiple sequence alignment of mature peptides of FMRFamide genes from various species belonging to Phylum Arthropod. Amino acid R at 2nd last position near to C terminal end is highlighted in blue color as R. The color distinction represents the Class distinction of species to which they belong, blue is for Class for Hexapoda, yellow is for Class for Crustacea and green is for Class Chelicerata. The consensus sequence is given at the end of the alignment.

(6)

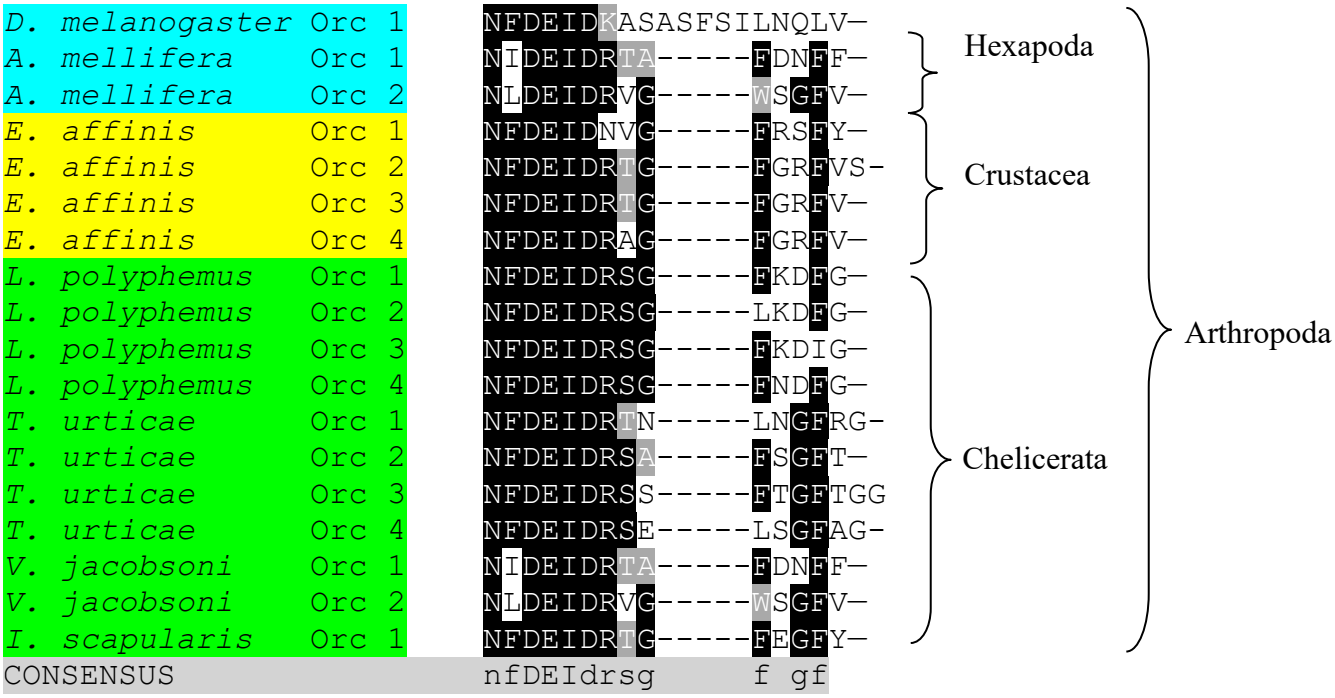


Figure 2.6. Multiple sequence alignment of orcokinin B mature peptides of from orthologus Orcokinin genes across the Phylum Arthopoda.

(7a)

Varroa destructor

```
MERPLLSFEGLIFLLLLLLYQIKPIHSDPAASESGGSGNSMVAQLLAGKSGIEESALREFDEEAYDVCRRYIALHDT
ERRTGVHLPLQGGNFRESVGGGSLNTARDGIVGDEEDGEATSTGIALSREIRRPPSRARVYAYRGQGEEGDYGVSTE
YTKVVSSKKRSCIRRGGSCDARPGDCCYHSSCRCNLWGTNCRCMRMGLLQRWINGRRR
```

Varroa jacobsoni

```
MERPLLSFEGLIFLLLLLLYQIKPIHSDPAASESGGSGNSMVAQLLAGKSGIEESALREFDEEAYDVCRRYIALHDT
ERRTGVHLPLQGGNFRESVGGGSLNTARDGIVGDEEDGEATSTGIALSREIRRPPSRARVYAYRGQGEEGDYGVSTE
YTKVVSSKKRSCIRRGGSCDARPGDCCYHSSCRCNLWGTNCRCMRMGLLQRWINGRRR
```

(7b)

(7c)

Varroa destructor

```
TCAAGTAGCTATAAGCAAGTTTGCTTGGAGCGCACACTTGGTGAGGAAGACAGCGACGACAGGCAGTTTTATGCAAAGGACGGCAA
AACAAAGCGAGTTGCAGTAGAAGCGTACGGCACTCAGCTTTGCTGCAAGGGGCCGAACGGAAACAGCGTACGCAGAAATTTGAGTG
AATCACATTGAACACAGTAATTACTTGCTACTAGTCTTGCCTAATCCCTTCCCGCCTTTCTTCGCTCCTGCAGAACAGACTAAACT
TATTTTACTCAACAACACTGGTAACTGCATTATCGAAATTTTTAGCAGAAAGAGTTGACATTGTTATTAAGAACAAAGTAAAGGTG
ACTAATCTCCTTATTCACTAACTGCTAATTAATGAACGACTATTACATTATTAAGGTGTGCAAAAGTGTTGGTGCCAGAGGGTGAA
CGACTAATAAAGAAGAATCTCAGTGAAGACGAAACAGTTTCTGTGCGATTCTGACAAAGTATTATCGGACCGTATTGTGTTGTCTG
GTGAAACTGAAAGCTGAAGACGGCAAAGGCTGGCAGTCGGCTTCGCTTAAAAGCGGGAAAAATCTCGTCTACTCGTGTGAGCGAGC
TGCGTGTCAGCTCGCGATTGTTCCTGCTACGTCGTGGCATGGAGAGGCCCCTGTTATCCTTCGAAGGACTCATCTTCCTGCTATTG
                                   M  E  R  P  L  L  S  F  E  G  L  I  F  L  L  L
TTACTACTCTATCAAATTAAGCCCATTCACTCAGATCCGGCGGCATCGGAAAGTGGTGGAAGCGGAAATTCTATGGTGGCTCAGCT
 L  L  L  Y  Q  I  K  P  I  H  S  D  P  A  A  S  E  S  G  G  S  G  N  S  M  V  A  Q  L
GTTGGCTGGAAAAAGCGGGATTGAGGAATCGGCATTACGAGAGTTCGATGAAGAAGCATATGACGTTTGCCGACGGTATATTGCAC
  L  A  G  K  S  G  I  E  E  S  A  L  R  E  F  D  E  E  A  Y  D  V  C  R  R  Y  I  A
TGCACGATACCGAACGACGCACTGGAGTGCATCTCCCACTACAAGGTGGGAACTTTCGGGAATCGGTTGGCGGTGGGTCTTTAAAC
L  H  D  T  E  R  R  T  G  V  H  L  P  L  Q  G  G  N  F  R  E  S  V  G  G  G  S  L  N
ACTGCAAGAGATGGTATTGTAGGAGACGAGGAAGACGGTGAAGCGACTTCGACAGGAATTGCCCTTTCGAGGGAGATCAGGCGGCC
 T  A  R  D  G  I  V  G  D  E  E  D  G  E  A  T  S  T  G  I  A  L  S  R  E  I  R  R  P
GCCATCACGAGCCCGCGTTTATGCCTACAGAGGGCAAGGCGAAGAGGGCGACTACGGAGTATCGACAGAGTACACAAAAGTGGTCT
  P  S  R  A  R  V  Y  A  Y  R  G  Q  G  E  E  G  D  Y  G  V  S  T  E  Y  T  K  V  V
CTTCGAAAAAGCGCAGCTGTATCAGACGTGGCGGATCTTGCGACGCTCGTCCGGGCGACTGCTGCTATCATTCGTCATGTCGCTGC
S  S  K  K  R  S  C  I  R  R  G  G  S  C  D  A  R  P  G  D  C  C  Y  H  S  S  C  R  C
AATCTTTGGGGCACCAACTGCCGATGCATGCGTATGGGCCTACTGCAGAGATGGATCAATGGTCGGCGTCGCTGAATGCCAGATAT
N  L  W  G  T  N  C  R  C  M  R  M  G  L  L  Q  R  W  I  N  G  R  R  R  *
```

GCAGAGAACGCGGTTACGACCAGCTGCTGACACCCGCTACCCCAGTGATTACAGCTCTTACTACTACTACTATTATGATAACACTA
CGCACTATGCTGATGCCGCTGCTCGAGATTAAATCGAATATATGATACGATAACAATGACATTGCAGATCCGCATGTCATATAATC
GTACCAAGGTGACTTGTATCAAACGATTACCTAATTCGTTACTATTGTAATAAGCGAAGTGACGCAGCGAATCGGATATATGAGCG
ACGGTGGCATTCTTGCAATTACCGCAAACCCCCCACCCCACTCCCGCCTCTACCTGACGCCTACGCTACGAAAGTGAAAAATTCAG
AGCCCCAATAAATACGTACTGCATCATCATTCGAAATGTTGTCAGACTAGACTTTATCTAAAAGCACTAATGAAACTATAACAAGC
AAAATATTTATAGGTATGACACGTTTGCATTTATAGACGGAACGCTGGTTAGCGCTTCATATGATCCGACGATCCTTAACGAAACA
TCTTGAAATTCTAGGTCTAGGAAGAAAGGTCACGTTCCTTAGTACAGCTTGTGTTCTCGTTCGCGCGTGCGACGACCTTACTAGCA
CATCGAATGCAATGTGCTTCAAGCCACTCCAGAAGAAGAGAAGAACCTAGACAAAGACCAGTAGAATCCTCATTATAGCTTCGTCG
TCTTAGATGACCCAATAGCTTAGGTCCTGATCTCTTCGAAATCAAGTTGTTTCTCTCGAAGCTCGGTGCTTTCTTCTTAAAGCAAA
CATGCTTTTACGGTTGTACGTAATCTGTCCGCCCGAATCAGTTAAATGCAATTCAAAGAGCTCGCTTAAATAGAAGAGAGAAAAAT
TTCTTATCACCTATCGGCTAAATGCACATTATAATGCCATCCAACGGGTGGCGATATAGCGAAGAAAAAATACAGGGACATTTGCG
CAACAATCCTCACTACCCGTACGTGTTTTCCGTCGGTTGAAGACATAAAAAACGGAAATTTATCATATGTCTTTCTTAAGCACGTT
CTCTAATTGACATATTAGGTCGACGCAACTCGTGTTGCGTATATTTTGGTCACTTGGGAACCTGAGCGGGTCCCTCGCTGACAGAC
TTCTCAGAATGACTTCGCTTAATGACCGCTTCTGTCTCCTTTGCCTAGCTTCTTCGCTTCAATTCTTTTCAACAGCTACGGCTTCG
ACGCTTCATTGGATTCGCCATGTCCCTCTGGTCTACCTCATTCTTGATTACCCAAGTGGCGAACACTCGTTTAGCCAGAAGCAGCT
TGTCATGACATGGCAACATTTGCTGCTACTGAACCTCGTGGTAGGTATAACTATGACTGAATTTTTACGAATATTAGTAGTGATTA
AGGATCATTGCTGTTGCCATATAACAGTACTAACAATAGCTAATATGACTATTAGTATTTTGCCGACCTAGCGTCGCCAAGCGTCG
GCGATTCATGGAACAGTCCGAGGAACCTCTCGCACGCAAAAGTACTCTAGTAAATATCTCATTGATAGAGAGTAAAAACGAGTAGC
AATCTAACACAGTCTGATCAATGTAGCCTGATTATTATTGTTCAGATTCGTGGGTTTCGGATACGTAGCTCTGATACGCGAAGGCC
CCGTTGCTCTCGCCGTTCCAATGTTGAAGTGCTAGTCGTGATCAATTGAACAATAAACAAATGAATAATTAAATCGAATAAATATT
GAATACATCAAGTATGTCAGCTGTCTATGTCAATGTGATTAGTTCTTTTTCCTCAGAAATACTTATATATATATGCAGAAGAAAAT
ATATATATATAAGCAAGTACGCCCTGACTAGTAGATACAGAACTGCGTGTCGGTGATATACAGGTGTCATAGCACAGGGTAATATC
CTTTATCTATCCCGTAACAGGATCGCTTTGGTTTATACTTAACTGTAGTGA

Varroa jacobsoni:

GTGTGAGCGAGCTGCGTGTCAGCTCGCGATTGTTCCTGCTACGTCGTGGC**ATG**GAGAGGCCCCTGTTATCCTTCGAA
M E R P L L S F E

GGACTCATCTTCCTGCTATTGTTACTACTCTATCAAATTAAGCCCATTCACTCAGATCCGGCGGCATCGGAAAGTGG
G L I F L L L L L L Y Q I K P I H S D P A A S E S G
TGGAAGCGGAAATTCTATGGTGGCTCAGCTGTTGGCTGGAAAAAGCGGGATTGAGGAATCGGCATTACGAGAGTTCG
G S G N S M V A Q L L A G K S G I E E S A L R E F
ATGAAGAAGCATATGACGTTTGCCGACGGTATATTGCACTGCACGATACCGAACGACGCACTGGAGTGCATCTCCCA
D E E A Y D V C R R Y I A L H D T E R R T G V H L P

CTACAAGGTGGGAACTTTCGGGAATCGGTTGGCGGTGGGTCTTTAAACACTGCAAGAGATGGTATTGTAGGAGACGA
L Q G G N F R E S V G G G S L N T A R D G I V G D E
GGAAGACGGTGAAGCGACTTCGACAGGAATTGCCCTTTCGAGGGAGATCAGGCGGCCGCCATCACGAGCCCGCGTTT
E D G E A T S T G I A L S R E I R R P P S R A R V

ATGCCTACAGAGGGCAAGGCGAAGAGGGCGACTACGGAGTATCGACAGAGTACACAAAAGTGGTCTCTTCGAAAAAG
Y A Y R G Q G E E G D Y G V S T E Y T K V V S S K K
CGCAGCTGTATCAGACGTGGCGGATCTTGCGACGCTCGTCCGGGCGACTGCTGCTATCATTCGTCATGTCGCTGCAA
R S C I R R G G S C D A R P G D C C Y H S S C R C N

TCTTTGGGGCACCAACTGCCGATGCATGCGTATGGGCCTACTGCAGAGATGGATCAATGGTCGGCGTCGC**TGA**
L W G T N C R C M R M G L L Q R W I N G R R R *
ATGCCAGATATGCAGAGAACGCGGTTACGACCAGCTGCTGACACCCGCTACCCCAGTGATTACAGCTCTTACTACAACTACTATTA
TGATAACACTACGCACTATGCTGATGCCGCTGCTCGAGATTAAATCGAATATATGATACGATAACAATGACATTGCAGATCCGCAT
GTCATATAATCGTACCAAGGTGACTTGTGTCAAACGATTACCTAATTCGTTACTATTGTAATAAGCGAAGTGACGCAGCGAATCGG
ATATATGAGCGACGGTGGCATTCTTGCAATTACCGCAAACCCCCCACCCCACTCCCGCCTCTACCTGACGCCTACGCTACGAAAGT
GAAAAATTCAGAGCCCCAATAAATACGTACTGCATCATCATTCGAAATGTTGTCAGACTAGACTTTATCTAAAAGCACTAATGAAA
CTATAACAAGCAAAATATTTATAGGTATGACACGTTTGCATTTATAGACGGAACGCTGGTTAGCACTTCATATGATCCGACGATCC
TTAACGAAACATCTTGAAATTCTAGGTCTAGGAAGAAAGGTCACGTTCCTTAGTACAGCTTGTGTTCTCGTTCGCGCGTGCGACGA
CCTTACTAGCACATCGAATGCAATGTGCTTCAAGCCACTCCAGAAGAAGAGAAGAACCTAGACAAAGACCAGTAGAATCCTCATTA
TAGCTTCGTCGTCTTAGATGACCCAATAGCTTAGGTCCTGATCTCTTCGAAATCAAGTTGTTTCTCTCGAAGCTCGGTGCTTTCTT
CTTAAAGCCAACATGCTTTTACGGTTGTACGTAATCTGTCCGCCCGAATCAGTTAAATGCAATTCAAAGAGCTCGCTTAAATAGAA
GAGAGAAAAATTTCTTATCACCTATCGGCTAAATGCACATTATAATGCCATCCAACGGGTGGCGATATAGCGAAGAAAAAATACAG
GGACATTTGCGCAACAATCCTCACTACCCGTACGCGTTTTCCGTCGGTTGAAGACATAAAAAACGGAAATTTATCATATGTCTTTC

TTAAGCACGTTCTCTAATTGACATATTAGGTCGACGCAACTCGTGTTGCGTATATTTTGGTCAGTTGGGAACCTGAGCGGGTCCCT
CGCTGACAGACTTCTCAGAATGACTTCGCTTAATGACCGCTTCTGTCTCCTTCGCCTAGCTTCTTCGCTTCAATTCTTTTCAACAG
CTACGGCTTCGACGCTTCATTGGATTCGCCATGTCCCTCTGGTCTACCTCATTCTTGATTACCCAAGTGGCGAACACTCGTTTAGC
CAGAAGCAGCTTGTCATGACATGGCAACATTTGCTGCTACTGAACCTCGTGGTAGGTATAACTATGACTGAATTTTTACGAATATT
AGTAGTGATTAAGGATCATTGCTGTTGCCATATAACAGTACTAACAATAGCTAATATGATTATTAGTATTTTGCCGACCTGTCGCC
AAGCGTCGGCGATTCATGGAACAGTCCGAGGAACCTCTCGCACGCAAAAGTACTCTAGTAAATATCTCATTGATAGAGAGTAAAAA
TGAGAAGCAATCTAACACAGTCTGATCAATGTAGCCTGATTATTATTGTTCAGATTCGTGGGTTTCGGATACGTAGCTCTGATACG
CGAAGGCCCCGTTGCTCTCACCGTTCCAATGTTGAAGTGCTAGTCGTGATCAATTGAACAATAAACAAATGATTAATTAAAACGAA
TAAATATTGAATACATCAAGTATGTCAGCTGTCTATGTCAATGTGATTAGTTCTTTTTCCTCAGAGATGCAGAAGAAAATATATAT
ATATATAAGCAAGTACGCCCTGACTAGTAGATACAGAACTGCGTGTCGGTGATATACAGGTGTCATAGCACAGGGTAATATCCCTT
ATCTATCCCGTAACAGGATCGCTTTGGTTTATACTTAACTGTAGTGAAAGGAAGCTTTAAATGAATGTCTATCCATAGCGTCGCGG
TGAGTCTATCACCAGTCGTAATAGCGAAGTCTACGAACGGTCGCGTAAGTTTTAACTGTCTCCACACGAAACACAGCAGCCTGGTT
TACCTGGCGCAACATAGAGAGCTTGTCGGCACTACTGCCCTGTTCTATTATTCGATACCCGCTTAGGTCGTCTAAGTAAATCAGAG
TTGATAACGCCATCGAGGGAGGGACAAGGCGTTTAGGTTTCAGCTGACACTAATCCTAGGTTATATGTATCATTCTCTCGTGTAAA
TTTTAATGTATAATAAAGCAGTATATATTTATACAA

Figure 2.7 (A) Preprohormone gene sequences of agatoxin like neuropeptide in *Varroa* mite . The color scheme is same as indicated above: green color = signal peptide, yellow color = mature peptides, red color = canonical monobasic and dibasic cleavage sites. The sequence with red font represents accessory proteins. (B) Multiple sequence alignment mature peptide repeats, representing the conservation of the amino acid at particular position in respect with C terminal. The color distinction represents the Class distinction of species to which they belong, blue is for Class for Hexapoda, yellow is for Class for Crustacea and green is for Class Chelicerata. The consensus sequence is given at the end of the alignment. (C) Figure represents the gene structure for both *V. destructor* and *V. jacobsoni*. The signal peptide is indicated in bold and italics font. Red color represents the canonical dibasic cleavage sites, yellow color of sequence represents the mature peptide, pink color represents amidation site. Arrow represents the positioning of introns.

(8a)

Varroa destructor

MAVLLSVFIATVLCLAFLAGSAAGAEEQEHAFQGTATGHGRAASDTEMMNAIRTILLRSLRSPRYHGGTS
RRYDYVKRFPEVNARGFESDIFDEGFGDFSPVRRR

Varroa jacobsoni

MAVLLSVFIATVLCLAFLAGSAAGAEEQEHAFQGTATGHGRAASDTEMMNAIRTILLRSLRSPRYHGGTS
RRYDYVKRFPEVNARGFESDIFDEGFGDFSPVRRR

(8b)

(8c)

Varroa destructor

TCAAAGATGACAACTGCAACAAACGAACGTAAACAATAATGGCGCATAATTTTGAATAAGATCAAAAACGATGGTTG

AGTTCGACAAGTCGAAGTTTGTTGTGTTGCATTTCGTATTGAATGATCTAGCATAGTATGC**ATG**GCCGTGTTACTCA

M A V L L

GCGTGTTTATTGCTACGGTTTTGTGCCTAGCGTTTCTGGCCGGGTCAGCTGCCGGTGCCGAAGAACAGGAACATGCT

S V F I A T V L C L A F L A G S A A G A E E Q E H A

TTTCAGGGAACGGCAACTGGTCACGGAAGGGCTGCATCGGATACTGAGATGATGAATGCGATCAGAACGATTCTATT

F Q G T A T G H G R A A S D T E M M N A I R T I L L

GCGTTCTCTACGTTCTCCTCGCTACCATGGTGGTACCAGCCGCAGATACGATTATGTCAAGCGATTTCCGGAAGTCA

R S L R S P R Y H G G T S R R Y D Y V K R F P E V

ACGCCCGGGGCTTCGAGTCAGACATTTTTGACGAGGGCTTTGGTGATTTTAGCCCCGTCAGGCGCCGA**TAG**AGTAAT

N A R G F E S D I F D E G F G D F S P V R R R *

CATGCGACCACACAACAGAAATCCTCGAAGCCTGCGAAAGCGGCATTGAATGACAGGGTGGATGACGGGTGAATGAT
TGATGATCAAACGCTAATAATACAGTTATCGCCCGAACGGCAGGTAATTTATTGCACCCCGGAGGATGGAGCGAACA
GAAGATACCAGCTCGTGTATATTCACGCCCCAGTGAGAATCCAGCCTCCGGTCTCCGGGAAAATAGCACCAAAGGAA
AGGCAAAAACAGTTCCTAAAAAAATACCTCATCGTAACCTGTTGTGCTGAAAAGTTACGCCATAGACTTTGCGCGCA
GCAAGGATGCTTTCGCGGCAACACGTCATAGTGTCTAGTCAACATGTATTAACACGACGAACATAGCTGACGTCGCC
ATTGAAACGTGTAATGGAAAAAGATCATTCGTATAAGCACACGTGATCGCGTCACACTTTAGACTGATTTTACACTG
ATATAGATTTCTAATCAAATAGAACGATGTTACCATGCAAACAACAGCAGTTCGTGGATTTCCACCAGACCCCACTA
GGTACTGAATAATGACAGAAACGTCGTAACGGCAACGAGAGATCAACGGCAAAATGAAAATCAGCCCCACCAGTTAT
TATGAAGTGAAAATATCGATATCGTATCTCAGAAACTGAATCTCCCAAATTGATATCGCATAACTTACGACACGACA
CGACACGTTCCCACCGGAACGTTGCGCATAACTTAAGTGAATTACACTTACGCCGCGCTCGCAGCGACACACATGAC
ACGTTTACATGAAAACAGCATATTGAAAATTATAACTTATTCAAAACGACATATGTATATATATATACATATATTTC
TTTGATGCAACCAAAGAGAGGCGTGAATATTTGTTCGACGTTAATTATAAGACCCGAATGAGTGTTGGACTTCGCTG

AATTGATCACTGCGAAGGTTAGCGAACCAAAATTAGCTAAGAAAAATTGAAGCTTGTTAGTGTCGGGAACTATCCCC
CGATAGAGAACAACAATCATAGAAGACCTATCAATCAGTCAGCGAGAATGAAACGAACCCACATACACATGTCCCGC
ATAACCAACCCCTGAACATGAACCCCTGAACCCCTGAATCTCAAAACTGAAAGACTTGCTGGAAGAAACGCCACTAT
ACATTAGATAAGAAAAAAGTGACCGGATAAGACTATGTTAGTAATCCTATTTTCGAATATTAGCCAAATAAAAAATA
TATGCATCGCAGATTGTGTATATATATATA

Varroa jacobsoni

CCGAAGGGCATTGCTTAAAGCTGTACCAAAACTGCAATTAAATCTACGTGTTTGTGAAACCCATATTATATATATGG
GCTTATTATATATAACTTATTATAATATATAAGTTATGTATATATATATATATATATTAAGTTTACGGTTAGTGTTG
TTTTATTGGTGAATCTTTTCACCCTTGGACCAGCCCCTTTATAGAAGACGAAAACGGTAGATTCGACTAAATTATAG

TGTTAGATACTAGAACTCACGGATGTGATGGACAATGGCACAAGCCGAAATAGCCGTGTTACTCAGCGTGTTTATTG
M A Q A E I A V L L S V F I

CTACGGTTTTGTGCCTAGCGTTTCTGGCCGGGTCAGCTGCCGGTGCCGAAGAACAGGAACATGCTTTTCAGGGAACG
A T V L C L A F L A G S A A G A E E Q E H A F Q G T

GCAACTGGTCACGGAAGGGCTGCATCGGATACTGAGATGATGAATGCGATCAGAACGATTCTATTGCGTTCTCTACG
A T G H G R A A S D T E M M N A I R T I L L R S L R

TTCTCCTCGCTACCATGGTGGTACCAGCCGCAGATACGATTATGTCAAGCGATTTCCGGAAGTCAACGCCCGGGGCT
S P R Y H G G T S R R Y D Y V K R F P E V N A R G
TCGAGTCAGACATTTTTGACGAGGGCTTTGGTGATTTTAGCCCCGTCAGGCGCCGATAG
F E S D I F D E G F G D F S P V R R R *

AGTAATCATGCGACCACACAACAGAAATCCTCGAAGCCTGCGAAAGCGGCATTGAATGACAGGGTGGATGACGGGTGAA
TGATTGATGATCAAACGCTAATAATACAGTTATCGCCCGAACGGCAGAGGATGGAGCGAACAGAAGATACCAGCTCGTGTATATTC
ACGCCCCAGTGAGAATCCAGCCTCCGGTCTCCGGGAAAATAGCACCAAAGGAAAGGCAAAAACAGTTCCTAAAAAAATACCTCATC
GTAACCTGTTGTGCTGAAAAGTTACGCCATAGACTTTGCGCGCAGCAAGGATGCTTTCGCGGCAACACGTCATAGTGTCTAGTCAA
CATGTATTAACACGACGAACATAGCTGACGTCGCCATTGAAACGTGTAATGGAAAAAGATCATTCGTATAAGCACACGTGATCGCG
TCACACTTTAGACTGATTTTACACTGATATAGATTTCTAATCAAATAGAACGATGTTACCATGCAAACAACAGCAGTTCGTGGATT
TCCACCAGACCCCACTAGGTACTGAATAATGACAGAAACGTCGTAACGGCAACGAGAGATCAACGGCAAAATGAAAATCAGCCCCA
CCAGTTATTATGAAGTGAAAATATCGATATCGTATCTCAGAAACTGAATCTCCCAAATTGATATCGCATAACTTACGACACGACAC
GACACGTTCCCACCGGAACGTTGCGCATAACTTAAGTGAATTACACTTACGCCGCGCTCGCAGCGACACACATGACACGTTTACAT
GAAAACAGCATATTGAAAATTATAACTTATTCAAAACGACATATGTATATATATATACATATATTTCTTTGATGCAACCAAAGAGA
CGCGTGAATATTTGTTCGACGTTAATTATAAGACCCGAATGAGTGTTGGACTTCGCTGAATTGATCACTGCGAAGGTTAGCGAACC
AAAATTAGCTAAGAAAAATTGAAGCTTGTTAGTGTCGGGAACTATCCCCCGATAGAGAACAACAATCATAGAAGACCTATCAATCA
GTCAGCGAGAATGAAACGAACCCACATACACATGTCCCGCATAACCAGCCCCTGAATATGCGAATCTCAAAACTGAAAGAAACGCC
ACTATACATTAGATAAGAAAAAAGTGACCGGATAAGACTATGTTAGTAATCCTATTTTCGAATATTAGCCAAATAAAAAATATATG
CATCGCA

Figure 2.8 (A) Preprohormone gene sequences of Carausious like neuropeptide in *Varroa* mite . The color scheme is same as indicated above: green color = signal peptide, yellow color = mature peptides, red color = canonical monobasic and dibasic cleavage sites. The sequence with red font represents accessory proteins. (B) Multiple sequence alignment mature peptide repeats, representing the conservation of the amino acid at particular position in respect with C terminal. The color distinction represents the Class distinction of species to which they belong, blue is for Class for Hexapoda, yellow is for Class for Crustacea and green is for Class Chelicerata. The consensus sequence is given at the end of the alignment. (C) Figure represents the gene structure for both *V. destructor* and *V. jacobsoni*. The signal peptide is indicated in bold and italics font. Red color represents the canonical dibasic cleavage sites, yellow color of sequence represents the mature peptide, pink color represents amidation site. Arrow represents the positioning of introns

Table 2.1 Master table for neuropeptide genes sequence in *Varroa* mite

	Varroa destructor				*Varroa jacobsoni*			
Neuropeptide	**Gene ID**	**Genomic scaffold**	**Genome Location**	**Accession number**	**Gene ID**	**Genomic scaffold**	**Genome location**	**Accession number**
ACP	LOC111255317	NW_019211454.1	33898821..33953892	ND	LOC111255317	NW_019213753.1	49,643..49,831	ND
Agatoxin 1	LOC111251620	NW_019211458.1	6569367..6622879	XP_022664089.1	LOC111272178	NW_019214682.1	15,233..68,395	XP_022709201.1
Ast A	LOC111245583	NW_019211455.1	12812683..12847710	XP_022649853.1	LOC111266071	NW_019213949.1	440,750..471,460	XP_022698967.1
Ast B, MIP	LOC111251198	NW_019211458.1	13566989..13575746	XP_022663304.1	LOC111271347	NW_019214485.1	12947..22791	XP_022707817.1
Ast CC	NA	NW_019211459.1	457473...476208	NA	LOC111262149	NW_019212968.17	155715..174947	XP_022691927.1
AVPL	LOC111251516	NW_019211458.1	37691928..37694843	XP_022663895.1	LOC111260306	NW_019213330.1	55380..60593	XP_022688676.1
Bursicon alpha subunit	LOC111247436	NW_019211454.1	29962554..2996675	ND	LOC111264753	NW_019213797.1	111470..123359	XP_022696615.1
Bursicon beta subunit	LOC111264752	NW_019211454.1	106263..107475	XP_022663103.1	LOC111264752	NW_019213797.1	106263..107475	XP_022696614.1
CAPA	LOC111250488	NW_019211457.1	51714512..51760895	XP_022661568.1	LOC111259753	NW_019213266.1	31448..80768	ND
Carausious like peptide	LOC111243371	NW_019211454.1	21,911,437..21,917,170	XP_022644551.1	LOC111264254	NW_019213742.1	167,328..172,687	XP_022695704.1
CCHamide	LOC111251153	NW_019211458.1	41671372...41685058	ND	LOC111270805	NW_019213069.1	42877..79385	XP_022706958.1
CNMamide	ND	SRR8864012.46893032.2	ND	ND	ND	SRR3632582.185051089.2	ND	ND
Corazonin	LOC11125154	NW_019211458.1	14346111...14349321	ND	ND	NW_019213155.1	869,641..878,081	ND
CCAP	LOC111251821	NW_019211458.1	2879539...2886326	XP_022664588.1	LOC111263999	NW_019213714.1	212994..233585	ND
DH31	LOC111249899	NW_019213420.1	47101808..47111773	XP_022660097.1	ND	NW_019213420.1	40,143..41,446	ND

DH44	LOC111248063	NW_019211456.1	52194..65376	XP_022655526.1	LOC111262624	NW_019213586.1	82282..133025	XP_022692751.1
EFLamide	LOC111245583	NW_019211455.1	12812683..12847710	XP_022649853.1	LOC111270453	NW_019214375.1	141959..173733	XP_022706425.1
Elevenin	NA	SRR8864012.149089260.1	ND	ND	NA	NW_019213477.1		ND
FMRFamide	LOC111249730	NW_019211457.1	17231306..17336819	XP_022659724.1	LOC111260277	NW_019213729.1	226133..329050	XP_022688648.1
GPB5	NA	SRR8100122.28924566.2	ND	ND	NA	SRR3632582.77147140.2	ND	ND
ILP	LOC111244424	NW_019211459.1	26300365..26356647	XP_022647275.1	LOC111262602	NW_019212976.1	419103..499376	ND
ITP	LOC111254306	NW_019211457.1	26240076..26241570	XP_022670738.1	LOC111272461	NW_019214796.1	22103..23654	ND
Limostatin	NA	SRR8864012.120933285.2	ND	ND	NA	SRR3635105.101669407.1	ND	ND
MS	ND	ND	ND	ND	NA	SRR3635105.57299974.2	ND	ND
Natalisin	LOC111251926	NW_019211458.1	25869634..25889653	XP_022664805.1	LOC111262523	NW_019213574.1	49234..69894	XP_022692575.1
NPLP	LOC111245611	NW_019211455.1	37797402..37826958	XP_022649937.1	LOC111261041	NW_019213415.1	4..29540	XP_022689983.1
NPF	NA	SRR5377265.26255811.2	ND	ND	NA	SRR3632582.185285521.2	ND	ND
Orcokinin A+B	ND	ND	ND	ND	NA	SRR3632582.136591231.1	ND	ND
PDF	NA	SRR8864012.106407301.1	ND	ND	NA	SRR11879882.65162670.2	ND	ND
Proctolin	LOC111244398	NW_019211455.1	37292708…37,307,956	ND	NA	NW_019213315.1	156,016..156,126	ND
SIFamide	LOC111250327	NW_019211457.1	10421903..10431642	XP_022661115.1	LOC111264575	NW_019213779.1	51264..59285	XP_022696362.1
SK	LOC111244598	NW_019211455.1	22,115,403…22118584	ND	NA	NW_019214297.1	106235…107248	ND
Tachykinin	LOC111250114	NW_019211457.1	21471185..21539680	XP_022660581.1	LOC111265999	NW_019213941.1	149205..173384	XP_022698840.1
Trissin	NA	NW_019211456.1	4480771…4482045	NA	LOC111258892	NW_019213177.1	4479103-4484151	ND

- **Cells highlighted with blue color represents data entries from SRA database for partial and full length genes.**

- **Neuropeptide gene highlighted with orange color represents gene found only in one mite species.**
- **Neuropeptide genes highlighted with green color represent correction in the annotation at NCBI.**
- **Manual annotation was done for the genes such as corazonin, diuretic hormone 31 and neuropeptide F, for which accession numbers from SRA database is given.**
- **NA stands for not annotated while ND means not defined.**

Neuropeptide genes not found

- AKH
- Ast C
- Allatotropin
- CRF like-DH (DH37/47)
- EH
- ETH
- GPA2
- Gonadulin/ILP 8
- Hansolin
- Insect kinin/leucokinin
- Neuroparsin A
- PaOGS36577
- PTTH
- iPTH
- PK/PBAN
- Relaxin
- RFLamide
- RYamide
- sNPF

Table 2.2 Neuropeptide sequences among various orthologous species.

	Gene ID		Accession number			
Neuropeptide	***Varroa destructor***	***Varroa jacobsoni***	***Drosophila melanogaster***	***Apis Melifera***	***Ixodes scapularis***	***Tetranychus utricae***
ACP	LOC111255317	LOC111255317	ND	ND	ND	ND
Agatoxin like peptide	LOC111251620	LOC111272178		XP_003249808.1	XP_029835567.1	XP_025016647.1
AKH	ND	ND	NP_523918.1	XP_006563609.1	ND	ND
Ast A	LOC111245583	LOC111266071	NP_001287511.1	NP_001161181.1	XP_029848989.1	XP_015783579.1
Ast B, MIP	LOC111251198	LOC111271347	NP_648971.1	ND	XP_002434041.1	XP_015786537.1
Ast C	ND	ND	NP_523542.1	XP_006570429.1	XP_002433459.2	ND
Ast CC	SRR3635105.85904743.2	LOC111262149	NP_609483.2	XP_016907444.1	EEC00307.1	XP_015788542.1
Allatotropin	ND	ND	ND	ND	XP_002407036.2	ND
AVPL	LOC111251516	LOC111260306	ND	ND	XP_029841493.1	XP_015785540.1
Bursicon alpha subunit 2	LOC111247436	LOC111264753	NP_650983.1	XP_006570932.1	XP_002407512.1	XP_015794519.1
Bursicon beta subunit 5	LOC111264752	LOC111264752	NP_609712.1	NP_001035352.2	XP_002407513.2	XP_015784790.1
DH31	LOC111249899	SRR3632582.162967731.2	NP_723401.1	XP_026296832.1	ND	XP_015781297.1
CAPA	LOC111250488	LOC111259753	NP_524552.1	ND	XP_029825341.1	XP_015783322.1
Carausious like peptide	LOC111243371	LOC111264254	ND	ND	XP_029831523.1	LOC112539265
CCHamide	LOC111251153	LOC111270805	NP_001097784.1	XP_001120020.1	XP_029834159.1	ND
CNMamide	SRR8864012.46893032.2	SRR3632582.185051089.2	NP_001189021.1	XP_001121373.1	ND	ND
Corazonin	LOC 11125154	SRR8864012.127467345.2	NP_524350.1	NP_001012981.1	XP_029827221.1	ND

Crf like-DH (DH37/47)	ND	ND	NP_001097981.1	NP_001012633.1	EEC13046.1	XP_015785266.1
CCAP	LOC111251821	LOC111263999	NP_001262846.1	XP_001119829.1	XP_002402276.2	XP_015795188.1
DH44	LOC111248063	LOC111262624	NP_001097725.2	XP_026299101.1	EEC10289.1	XP_015788551.1
ETH	ND	ND	NP_524699.1	NP_001136079.1	ND	XP_015790707.1
EH	ND	ND	NP_001262668.1	XP_006557391.2	EEC02578.1	ND
Elevenin	SRR8864012.149089260.1	ND	ND	XP_003251276.1	AXL48134.1	XP_015781938.1
EFLamide	XP_022649853.1	XP_022706425.1	ND	ND	EEC18598.1	XP_015790582.1
FMRFamide	LOC111249730	LOC111260277	NP_523669.2	ACI90290.1	ND	XP_015790373.1
GPA2	ND	ND	NP_001104054.2	ND	XP_029847619.1	XP_015782352.1
GPB5	SRR8100122.28924566.2	SRR3632582.77147140.2	NP_001104335.1	ND	XP_029847632.1	XP_015781999.1
Gonadulin/ILP 8	ND	ND	NP_648949.2	XP_003250133.2	ND	ND
ILP	LOC111244424	LOC111262602	NP_648359.1	XP_026300673.1	EEC04486.1	XP_015782237.1
ITP	LOC111254306	LOC111272461	NP_001163294.1	XP_006571871.1	XP_029827278.1	XP_015785389.1
Insect kinin/leucokinin	ND	ND	NP_524893.2	XP_016768546.1	XP_029824648.1	XP_015790585.1
Limostatin	SRR8864012.120933285.2	SRR3635105.101669407.1	NP_611138.2	XP_001121053.2	ND	ND
MS	ND	SRR3635105.57299974.2	NP_001262911.1	ACI90289.1	ND	XP_015786756.1
Natalisin	LOC111251926	LOC111262523	NP_001097789.1	ND	XP_029837413.1	XP_015792419.1
NPLP	LOC111245611	LOC111261041	NP_001137760.1	XP_006559359.1	XP_029848113.1	ND
NPF	SRR5377265.26255811.2	SRR3632582.185285521.2	NP_001262642.1	XP_006559366.1	ND	XP_015786552.1
Neuroparsin A	ND	ND	ND	NP_001035359.1	XP_029849075.1	ND
Orcokinin A+B	ND	SRR3632582.136591231.1	NP_001261160.1	XP_016770158.1	XP_029842431.1	XP_015781985.1
PDF	SRR8864012.106407301.1	SRR11879882.65162670.2	NP_524517.1	XP_006570344.1	ND	ND
Proctolin	LOC111244398	SRR3635105.58662445.1	NP_001285728.1	ND	ND	XP_025016906.1

PTTH	ND	ND	NP_001303304.1	XP_555854.2	EEC00904.1	ND
PTH	ND	ND	ND	ND	XP_002411601.2	XP_025017087.1
Pyrokinin (PK)/PBAN	ND	ND	NP_524329.1	ABJ09601.1	XP_029825341.1	XP_015783322.1
RELAXIN/ILP 7	ND	ND	CCF74138.1	ND	ND	XP_015785150.1
sNPF	ND	ND	NP_724239.1	XP_003250155.1	ND	ND
SIFamide	LOC111250327	LOC111264575	NP_001246496.1	XP_003249680.1	ADD92393.1	ND
SK	LOC111244598	ND	NP_524845.2	XP_006557714.2	XP_029835307.1	ND
Tachykinin	LOC111250114	LOC111265999	NP_001262493.1	BAC76400.1	ND	XP_015792263.1
Trissin	ND	LOC111258892	NP_650471.2	XP_001121373.1	ND	ND

- **Cells highlighted with blue color represents data enteries from SRA database.**
- **Neuropeptide gene highlighted with orange color represents gene found only in one mite species.**
- **Neuropeptide genes highlighted with green color represent correction in the annotation at NCBI.**
- **Manual annotation was done for the genes such as corazonin, diuretic hormone 31 and neuropeptide F, for which accession numbers from SRA database are given.**

Chapter 3 - Annotation of *Varroa* mite neuropeptide G protein-coupled receptors

Introduction

Neuropeptides are small proteins that regulate the development, growth, reproduction, metabolism and behavior of arthropods (Nässel, 2018; Nusbaum et al., 2017). While the majority of neuropeptides used to modulate biological functions by interactions with G protein-coupled receptors (GPCRs, metabotropic), others can also interact with other types of receptors, including Receptor Tyrosine-Kinases (RTKs) (Vogel et al., 2015), Peptide-gated ion channels (FMRFamide-gated sodium channels, ionotropic) (Dürrnagel et al., 2010), and membrane-bound guanylate cyclase (GC) receptors (Chang et al., 2009). GPCRs, RTK and GC are found in insects, while peptide gated ion channel type neuropeptide receptors were found in more ancestral lineages of animal, snails and hydra (Dürrnagel et al., 2010), but not in arachnids yet.

GPCRs, integral membrane proteins with seven transmembrane domains, transduce the extracellular signals (chemicals/ligands) across the membrane via interaction with trimeric G-proteins. These membrane proteins are one of the most numerous and diverse protein families found in mammals (Kroeze et al., 2003Z). The genomes of bacteria, yeast, plants, nematodes, and other animals species all contain the genes encoding GPCRs. GPCRs are classified into five distinct classes: Class A (Rhodopsin type), Class B1 (Secretin), Class B2 (Adhesion type), Class C (Glutamate type), and Class F (Frizzled) (Attwood and Findlay, 1994; Kolakowski Jr, 1994; Hu et al., 2017). Class A, the largest GPCR subfamily, is primarily composed of receptors for olfactory, light (rhodopsin), and adrenaline-based ligands. Class B receptors include those for the secretin family of gastrointestinal/gut-derived peptide hormones, corticotropin-releasing

hormone, and calcitonin, and Class C receptors include those for metabotropic glutamate-based ligands (Pierce et al., 2002).

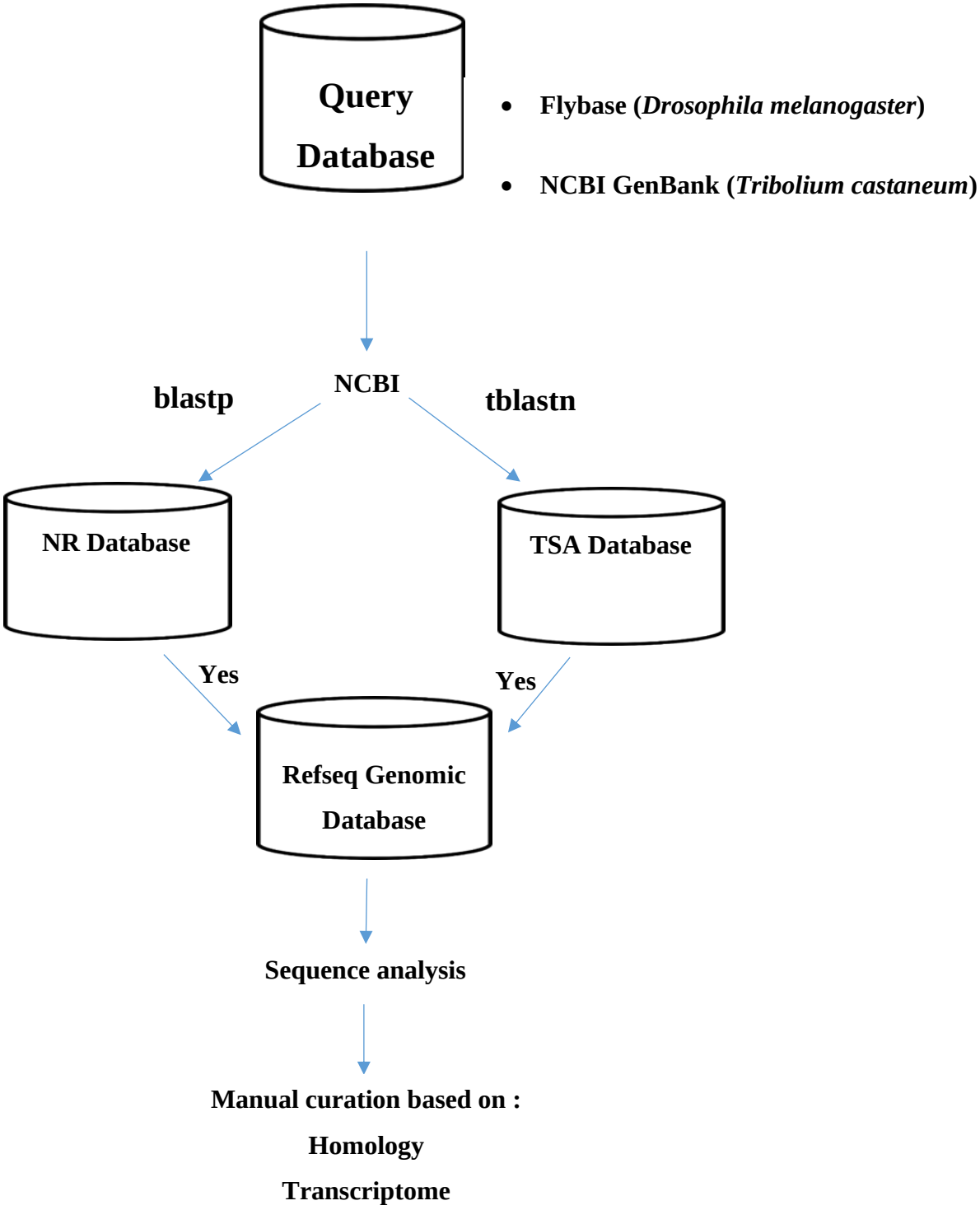
GPCRs regulate a broad range of physiological functions by reacting to and interacting with cellular environmental signals (light, biogenic amines, neuropeptides, peptide hormones, and protons). Typically, G-proteins are present in the trimeric form in the inactive state of the receptor. Once a ligand binds at the receptor site, activation of the receptor occurs through interaction with trimeric G-proteins (Gα, Gβ, and Gγ). The receptors are known to interact with the Gα subunit through TM3, TM5, TM 6, and intracellular loops 2 and 3 (ICL2 and ICL3) (Hilger et al., 2018). The Gα (alpha subunit) mediates the exchange of GDP (guanosine diphosphate) with GTP (guanosine triphosphate) and thereby activates the receptor. When the receptor is actived, the Gα dissociates from the Gβγ dimeric subunit. At the downstream level, both Gα and Gβγ regulate the activity of different proteins. The activation of the receptor is completed when GTP is hydrolyzed back to GDP, and the Gα reunites with Gβγ dimeric subunit.

In the history of GPCRs discovered in insects, the Tachykinin receptor from *D. melanogaster*, was the first functionally characterized neuropeptide GPCR (Li et al., 1991). This was followed by identification of only a few more GPCRs including the neuropeptide Y (NPY)-like receptor, gonadotropin releasing hormone like receptor, Allatostatin receptor (from *D. melanogaster*), and diuretic hormone receptor (from *Manduca sexta* and *Acheta domesticus*) (Birgül et al., 1999; Hauser et al., 1997; Li et al., 1992; Reagan, 1994, 1996). Further, as a result of technological advancements, a large set of genomic information of other insect species became available, providing valuable insights into the evolution of neuropeptide signaling across various insect orders and allowing for the deorphanisation of the receptors at comparatively faster pace. For example, in *D. melanogaster*, the gonadotropin like neuropeptide receptor was

deorphanized as adipokinetic hormone receptor (AKH) (Staubli et al., 2002) and NPY-like was deorphanized as RYamide receptor (Collin et al., 2011; Ida et al., 2011a). Additionally, newer neuropeptide signaling systems were discovered in different insects that were absent in *D. melanogaster* such as AKH/corazonin like peptide (ACP), allatotropin (AT) and inotocin were found in *Anopheles gambiae*, *Bombyx mori* and *Tribolium castaneum*, respectively (Hansen et al., 2010; Stafflinger et al., 2008; Yamanaka et al., 2008). Generally, the numbers of the GPCR for neuropeptide and protein hormone are between 35-50 in an insect species (Caers et al., 2012; Vogel et al., 2013).

This study was focused on finding the neuropeptide and cognate receptor genes from the genomic scaffold of the *Varroa* mite (*V. destructor* and *V. jacobsoni*) in order to gain a better understanding of how the neuropeptide signaling has evolved in the *Varroa* mite by comparing it with species from closely related taxa including Class Insecta, *D. melanogaster* and *A. mellifera*, and Class Chelicerata, *Ixodes scapularis* and *Tetranychus urticae*. Also, the findings of the study will help in identifying a neuropeptide signaling system that could be targeted for developing bee safe acaricides through comparative genomics approach.

Figure 3.1: Flowchart Diagram 2, Methodology



Initially, gene sequences of known GPCRs from other species (*Drosophila melanogaster*, *Apis mellifera*, *Ixodes scapularis* and *Tetranychus urticae*) were used to mine non redundant database of *Varroa* mite at National Centre for Biotechnology Information, NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) for potential putative receptor gene sequence. The blastp program was used and followed by the manual correctness of the search result for the reliable matches based on the high percentage sequence similarity and high blast e-value. Gene annotations on the putative translations were made for the putative transmembrane regions in the sequence.

When the search was unsuccessful in the nr database, the search was expanded to other databases at NCBI: RefSeq genome database and Transcriptome shotgun assembly database (TSA) using the tblastn program to look upon unannotated sequences. The search strategy was same as mentioned above. When the match was found in the RefSeq genome database, the gene prediction was made using the FGENESH program (http://www.softberry.com/) for the sequence surrounding the matching region; 1500-2000 kb of each side of the match was entered into the program to find the potential entire gene coding sequence.

The orthology relationships in the GPCR search result used two criteria: 1) back blast using the search output against *D. melanogaster* and 2) analysis of orthology via phylogenetic clustering using the neighborhood joining method. If no orthologs are found, the search is shifted to another database until the correct orthologous gene sequence is discovered. Additionally, the expected threshold value in the range of 10-1000 was adjusted for the purpose of determining the correct gene sequence.

The seven transmembrane regions of the GPCRs are generally conserved among orthologous genes. The gene structure, i.e. the intron and the exon splice sites in the receptor

sequence, was manually curated using the GT-AG rule; GT residues (GT = glycine and thiamine) at the 5' end of the intron donor at the splice site and AG residues (AG = adenine and guanine) at the 3' end of the intron acceptor. This rule is found in the 98% of eukaryotic mRNA processing cases (Alberts et al., 2013; Burset et al., 2001). The transmembrane regions in the curated protein sequence were predicted using the TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/)(Sonnhammer et al., 1998). Additionally, the putative receptor sequence was compared to that of its orthology to determine its conservation. To accomplish this, the receptor gene sequence was aligned with the sequences of orthology using the MEGA-X software's MUSCLE alignment (Edgar, 2004). The multiple sequence alignment was manually curated to ensure that the transmembrane regions of the multiple sequences were properly aligned. The multiple sequence alignment was sed to conduct phylogenetic analysis by clustering of the receptor gene sequences using a neighborhood joining program with a 1000 bootstrapping.

Results

In the current study, we identified about 27 cognate/putative GPCRs following the methodology mentioned above. However, neuropeptides such as Eclosion hormone and Prothoracicotropic hormone, as well as insulin-like peptides (Relaxin and gonadulin), are known to be acting through membrane guanyl cyclase, receptor tyrosine kinase, and leucine rich repeat GPCR, respectively. The putative receptors for these neuropeptides will be excluded from discussion. GPCRs belonging to Class A that were found are for the receptors: ACP, Allatostatin A, B and C, Bursicon, CCHamide, Corazonin, CCAP, Myosuppressin, Neuropeptide F, Proctolin, Pyrokinin, RYamide, Short Neuropeptide F, SIFamide, Sulfakinin, Tachykinin, Trissin. Neuropeptide GPCRs belonging to Class B that were found for the receptors: Diuretic

hormone 31 type (Calcitonin like) and 44 type (CRF like), Pigment dispersing factor (PDF) and insect Parathyroid hormone.

Class A type GPCRs

AKH related peptide receptors:

Adipokinetic hormone (AKH) and gonadotropin-releasing hormone (GnRH) are orthology groups found in insects and in vertebrates, respectively. Often it is also categorized with the name GnRH and related neuropeptide superfamily present in all bilaterian species (Mirabeau and Joly, 2013). GnRH ligands and their receptors are found in vertebrates, while GnRH like ligands including AKH, corazonin, and AKH-Corizon related peptide (ACP) are reported in arthropods (Zandawala et al., 2018). Gene duplication events have resulted in independent AKH, Corazonin, and ACP neuropeptide signaling in arthropods, where ACP signaling has been independently lost from a number of arthropods lineages (Hansen et al., 2010; Hauser and Grimmelikhuijzen, 2014). Although corazonin signaling is found in various protostomes and dueterostomes, it is lost in vertebrates, urochordates and nematodes, likely due to lineage specific losses (Roch et al., 2011).

AKH and its GPCRs play vital role in insects as these mobilize carbohydrates and lipids from fat bodies (in comparison to the liver and adipose tissues of mammals), in both during larval and adult stages (Hansen et al., 2010; Li et al., 2016). AKH signaling is crucial for insect locomotion, especially flight. However, AKH receptor was not found in the *Varroa* mite species. Other arachnid, the genome sequences of *I. scapularis*, also lack the AKH receptor. Additionally, a gene sequence encoding the AKH neuropeptide was not found in either of the *Varroa* mite species. Considering both neuropeptide and receptor results, absence of AKH type signaling is likely a true loss of the AKH signaling.

Apart from the AKH signaling, ACP, and corazonin signaling systems were identified in the genomic scaffold of *Varroa* mite species. Corazonin was initially isolated from the American cockroach, *Periplaneta americana,* for its cardio-accelerating properties. ACP neuropeptide signaling system was recently separated from AKH with the authentic independent receptor (Hansen et al., 2010). The ACP neuropeptide and GPCR receptor are structural intermediates between AKH and corazonin, and thus are referred to as the ACP neuropeptide and receptor. Despite being a structural intermediate between AKH and Corazonin, the ligand specificities have been established in the course of evolution. Based on sequence similarity, and phylogenetic clustering (neighbour joining tree method with 1000 bootstrapping tests) with other arthropod ACP and corazonin receptors, receptors for ACP and corazonin were found in the genomic scaffold of *Varroa* mite species (Figure 1). Seven transmembrane domains (TMs) were identified in the receptor sequence (Figure 3.8A and 3.9A), indicating that the gene prediction covers the entire open reading frame.

CCAP receptor :

Crustacean cardioactive peptide (CCAP) was first identified in the shoe crab, *Carcinus maenas* (Lee et al., 2013; Loi et al., 2001; Stangier et al., 1987), and is a conserved nonapeptide in arthropods . CCAP has been implicated in a variety of functions in *D. melanogaster* and *Manduca sexta,* including myotropic activity and controlling motor programs associated with the ecdysis process. CCAP's cognate receptor is a GPCR, previously identified in chelicerates such as *T. urticae* (Veenstra et al., 2012) and insects such as *D. melanogaster* and *A. mellifera* (Cazzamali et al., 2003; Loi et al., 2001). *V. destructor*'s putative receptor gene sequence for CCAP had a blast result with the e-value of 5e-124 and a percent identity of 48%. The potential receptor gene sequence discovered in *V. jacobsoni* had a blast result with the e-value of 2e-124

and a percent identity of 54%. Clustering the *Varroa* mite's putative gene sequence for CCAP receptors with known CCAP receptors from other species provided final confirmation. Seven transmembrane domains (TMs) were identified in the receptor sequence (Figure 3.10A), indicating that the gene prediction covered the entire open reading frame.

Sulfakinin receptor:

Sulfakinin is an arthropod neuropeptide that is involved in feeding and digestion (Zels et al., 2014). Sulfakinin share a substantial degree of structural and functional similarity with the mammalian gastrin and cholecystokinin peptides (Hauser et al., 2006). The ligand sulfakinin was found in the genomic scaffold of the *Varroa* mite species. Additionally, a gene sequence that encodes for putative sulfakinin receptor was found from the genomic scaffold of both *V. destructor* and *V. jacobsoni*. The sequence found in both *Varroa* mites had a blast result with the e-value of 3e-58, and identity percent 53%. Finally, the receptor gene sequence was confirmed by phylogenetic clustering with 80% bootstrapping support in neighborhood joining tree method (1000 bootstrapping) with other known sulfakinin receptor gene sequences from other arthropod species. Seven transmembrane domains (TMs) were found in the receptor gene sequence, indicating that the gene prediction covers the entire open reading frame.

RYamide receptor:

RYamide, also known as luqin-like neuropeptide, was first reported in the arthropod *Cancer borealis* (Li et al., 2003). The molluscan luqin like signaling has a clear orthologous relationship with the arthropod RYamide signaling (Jékely, 2013). In insects, the RYamide was described in *D. melanogaster* and *Tribolium castaneum*. While the RYamide neuropeptide was not found in the genomic scaffold of both *Varroa* mites, a cognate receptor sequence was discovered in the genomic scaffold of *V. destructor* and *V. jacobsoni.* The sequence had a blast

result with the e-value of 4e-91 and a percent identity of 50%. Seven transmembrane domains (TMs) were found, indicating that the gene prediction covers the entire open reading frame. According to the theory of receptor-ligand coevolution (Park et al., 2002), it is highly improbable that a RYamide neuropeptide would be absent while its cognate receptor is present. This could be for one of two reasons: either the neuropeptide sequence has diverged significantly from the conserved arthropod RYamide sequence, escaping from the current blast algorithm, or it is possible that of the genome assembly of *Varroa* mite is yet incomplete.

Tachykinin and related peptide receptor

Tachykinin and related peptides, i.e., Natalisin as well as their receptors from *V. destructor* have been recently reported (Genbank accession numbers are XP_022673516.1 and XP_022644461.1 for the Tachykinin and Natalisin neuropeptide receptor sequences, respectively) (Jiang et al., 2016). These previously published sequences were used as a query to retrieve the receptor sequence for Tachykinin and Natalisin from the genomic scaffold of *V. jacobsoni*. Each sequence for Tachykinin receptor and Natalisin receptor was identified with high percentages of the identities (96.71 percent and 100 percent, respectively) to the query sequence from *V. destructor* and had a blast e-value of 1e-104. Seven transmembrane domains (TMs) were found in the receptor sequence, indicating that the gene prediction covers the entire reading frame.

CCH Receptor:

CCHamide is an arthropod ortholog of excitatory peptide (EP) which is commonly found in annelids and mollusk species (Thiel et al., 2019). CCHamide was first reported in silkworm *Bombyx mori* (Roller et al., 2008) and has since been discovered from chelicerates and crustaceans (Christie, 2015; Hansen et al., 2011; Nguyen et al., 2018; Toullec et al., 2017;

Veenstra, 2016). CCHamide's cognate receptor is a GPCR that was discovered through the de-orphanisation process (Bauknecht and Jékely, 2015; Hansen et al., 2011). CCHamide signaling is involved in sensory perception, feeding, and control of insulin-like peptides, specifically in *Drosophila melanogaster*, as well as in control of feeding in other insects (Thiel et al., 2019). Along with the CCHamide neuropeptide, its putative cognate receptor was found from the genomes of both *Varroa* mites. The sequence found in *V. destructor* and *V. jacobsoni*, had a blast result with the e-value and percent identity of 5e-96, 8e-96, and 46.13%, 46.13%, respectively. The final confirmation was performed using phylogenetic clustering with 99% bootstrapping support in neighborhood joining tree (1000 bootstrapping) with other known CCHamide receptor sequences from other arthropod species. Seven transmembrane helixes (TMHs) were found in the receptor sequence (Figure 3.12A), indicating that the gene prediction was accurate.

SIFamide receptor:

SIFamide signaling was first reported from the Grey flesh fly, *Neobellieria bullata* (Verleyen et al., 2004) and has since been it has been reported from many arthropods. The SIFamide gene sequence has been reported to be highly conserved in terms of the mature peptide's location within the preprohormone sequence in various arthropods. SIFamide signaling has been associated with a variety of functions in arthropods species, including sexual behavior, sleep regulation, and pupal mortality in *D. melanogaster* (Sellami and Veenstra, 2015), and feeding control in *I. scapularis* via interacting with myoinhibitory peptide signaling (Šimo et al., 2009; Šimo et al., 2013). Two putative genes encoding the SIFamide receptors were found in the genomic database of both *Varroa* mites. The sequences found in *V. destructor* and *V. jacobsoni* had a blast result with the e-value and percent identity ranging between 5e-58 to 5e-63 and 38%-

41%, respectively. Seven TMs were found in the receptor sequence (Figure 3.13A), indicating that the gene prediction was accurate.

Neuropeptide F and Short neuropeptide F receptor:

Neuropeptide F (NPF) was first identified in tapeworms and now found in both protostome and deuterostome species (except vertebrates) (Maule et al., 1991; Nässel and Wegener, 2011). The first NPF receptor in other than mammalian species was identified in *D. melanogaster* (Garczynski et al., 2002). NPF is similar in sequence to vertebrate Neuropeptide Y peptide, except that it contains a F (Phenylalanine) at C-terminal end and has a consensus sequence of RPRFamide (Arg-Pro-Phe-amide). The mature peptide of NPF varies significantly in length, ranging from 8 to 34 amino acids. NPF signaling has been linked to feeding and reproduction regulation in species other than mammals. The gene sequence encoding the NPF neuropeptide was found in the genome of the *Varroa* mite. Additionally, two gene sequences encoding the NPF receptor were found in the genomic database of each *V. destructor* and *V. jacobsoni*. The Blast e-value and percent identity of the gene sequences discovered in *V. destructor* and *V. jacobsoni* range between 3e-98 to 3e-111 and 45% - 48%, respectively. In the last, receptor sequences (Figure 3.14A) were analyzed for the presence and location of TMs.

The short Neuropeptide F has a consensus sequence of M/T/L/FRFamide (Met/Thr/Leu/Phe-Arg-Phe-amide) (Fadda et al., 2019). The sNPF signaling system evolved before the evolutionary split into protostomes and deuterostomes, but no sequence encoding the sNPF neuropeptide gene was found in the genomic database of the *Varroa* mite. Surprisingly, two putative genes encoding the cognate sNPF receptor were found in each of *Varroa* mite species. This anomaly could be caused by one of two factors: either the neuropeptide sequence diverged significantly from the consensus sequence or the genomic assembly of *Varroa* mites is

yet incomplete. Seven TMs were found in the receptor sequence, indicating the full-length coverage of the open reading frame and gene prediction was correct.

Allatostatin A and C receptor:

In insects, allatostatin signaling is involved in the regulation of metamorphosis and reproduction. There are three neuropeptides within this signaling family: Ast-A, Ast-B, and Ast-C. Their cognate receptors are rhodopsin type Family A of GPCRs. The first functional characterization of allatostatin A neuropeptide signaling was made in the cockroach, *Diplopetera punctata* (Tobe and Bendena, 2013). The allatostatin A receptor is closely related to Galanin receptor found in vertebrates. In general, insects have a single gene encoding Ast-A receptor, whereas *T. castaneum* lacks a gene encoding Ast-A and the Ast-A receptor (Li et al., 2008). However, in flies and mosquitoes, two copies of the Ast-A receptor gene exist as a result of a gene duplication event. A single gene sequence encoding the Ast-A receptor was discovered in the genomic scaffolds of each *Varroa* mite speceis with a percent identity of 53%. When *D. melanogaster*'s Ast-A receptor was used as a query, the putative gene sequence for Ast-A receptor had a blast result with the e-value of 3e-86 and 8e-86, in *V. destructor* and *V. jacobsoni*, respectively. Seven TMs were found in the receptor sequence (Figure 3.15A), indicating the gene prediction was correct.

Allatostatin C has a characteristic PICSCF (Pro-Ile-Cys-Ser-Phe) consensus sequence at the C-terminal end of mature peptides. The Ast-C peptide, first reported in tobacco horn moths (Kramer et al., 1991; Veenstra, 2009a), is involved in a variety of functions that vary between species. It inhibits the JH hormone biosynthesis in some moths but instead regulates the heartbeat in *D. melanogaster*. Ast-C and its cognate receptor are consereved in arthropods. Single gene that encodes Ast-C receptor was found in the genomic scaffold in each *Varroa* mite

species. The sequence found had blast e-value and percent identity of 9e-128, 1e-127 and 59.27 % each in *V. destructor* and *V. jacobsoni,* respectively, when Ast-C receptor for *D. melanogaster* was used as a query.

PRXamide type neuropeptide receptors:

The peptides belonging to this group share a common C-terminal amide terminus, i.e., PRXamide (Pro-Arg-Leu/Met/Val/Ile-amide) (Jurenka, 2015). The cognate receptors for this group of peptides share a high degree of sequence homology. The PRXamide peptide family includes Pyrokinin – Diapause hormone and Pheromone Biosynthesis activating neuropeptide, CAPA (or CAP_{2b}), and Ecdysis triggering hormone. ETH is found in both protostomes and deuterostomes. Deuterostomes have evolutionarily related neuropeptide signaling as neuromedin-U. Indeed, the PBAN receptor in the moth *Helicoverpa zea* was discovered based on the similarity of the vertebrate neuromedin U receptor (Choi et al., 2003). Only nematodes, acari, and insects have been used to verify the presence of PRXamide receptors experimentally (Jurenka, 2015). In moth, *Helicoverpa zea*, the PBAN receptor was identified based on the sequence similarity between PRXamide receptors from *D. melanogaster* and vertebrate Neuromedin U receptor (Jurenka, 2015).

Among these neuropeptides, a putative sequence for the *capa* gene, which encodes the CAPA peptides, was identified in the genomic scaffold of each *Varroa* mite species. Additionally, one putative gene encoding the CAPA receptor was identified in both *Varroa* mites' genomic scaffold. The identified sequence in both mites had a percent identity and a blast e-value of around 40% and 6e-87, respectively. Surprisingly, one gene sequence was found in the genome of each *Varroa* mite species that showed similarity with the *D. melanogaster*'s Pyrokinin receptor gene sequence. The receptor gene sequence had a blast e-value of 6e-70 and a

percent identity of about 40%. Final confirmation was made by phylogenetic clustering with 57% bootstrapping support in neighborhood joining tree (1000 bootstrapping) the potential receptor sequence with known orthologous receptor gene sequences from closely related arthropods. In the last, sequence was analyzed for the presence and location of seven TMHs and were found, indicating that the sequence (Figure 3.17A) covers entire open reading frame.

Intriguingly, one Pyrokinin receptor sequence was identified, but no Pyrokinin neuropeptide gene sequence was identified in the *Varroa* sequence database. This anomaly could be caused by one of two things: either the neuropeptide sequence has diverged significantly from the consensus sequence, or the genome assembly of *Varroa* mites is yet incomplete.

ETH or ecdysis triggering hormone and its receptor gene sequences were not identified in the genomic database of both *Varroa* mites. However, ETH has been identified in other chelicerates including *I. scapularis* and *T. urticae* (Roller et al., 2010; Veenstra et al., 2012). Additionally, blast search revealed ETH neuropeptide and receptor gene in other chelicerates such as *Limulus polyphemus*, *Centruroides sculpturatus*, *Panonychus citri*. It is difficult to conclude that ETH signaling is suddenly lost in the *Varroa* mite, as some of the earlier inconclusive results suggests that the genomic assembly is yet incomplete. As a result, further confirmation is required before drawing any conclusions.

Proctolin receptor:

Proctolin is a pentapeptide with the consensus sequence RYLPT (Arg-Tyr-Leu-Pro-Thr). Proctolin has been isolated from various arthropods, including the horseshoe crab, *Limulidae* (Groome et al., 1990). Proctolin is the first neuropeptide in insects to be chemically characterized (Starratt and Brown, 1975). The sequence of Proctolin has a couple of variations, such as AYLPT (Ala-Tyr-Leu-Pro-Thr) and RYLMT (Arg-Tyr-Leu-Met-Thr), in *Leptinotarsa*

decemlineata and *Daphnia pulex*, respectively (Dircksen et al., 2011; Spittaels et al., 1995). Previous reports suggest the association of Proctolin with myotropic activities (Orchard et al., 2011). Proctolin neuropeptide and receptor gene sequences were identified from the genomic scaffold in both *Varroa* mites. The identified receptor sequences had a blast e-value and a percent identity of 5e-28, 3e-28, and 40.26 percent in *V. destructor* and *V. jacobsoni*, respectively.

FMRFamide receptor:

FMRFamide is originally a molluscan cardio acceleratory tetrapeptide. In general, arthropod FMRFamide peptides are a family of peptides with a C-terminal RF (Arg-Phe) amino acid motif (Coast and Schooley, 2011). FMRFamide related peptide family includes Myosuppressin, N-terminally extended FMRFamides, Neuropeptide F, and Short Neuropeptide F. N-terminally extended FMRFamides act as a modulator of gut and heart muscle contractions (Nagata, 2016). A gene that putatively encodes for the N-terminally extended FMRFamide neuropeptide was found, but its putative receptor gene was not found. Given that neuropeptides and their receptors co-evolve in arthropods (Li et al., 2008; Park et al., 2002), it is highly improbable that the neuropeptide gene is present in the genome, but the cognate receptor gene is absent. The receptor sequence is missing in both transcriptome data and genome assembly.

Myosuppressin Receptor:

Myosuppressin is a N-terminally extended decapeptide with a consensus sequence of XDXXHXFLRFamide (X-Asp-X-X-His-X-Phe-Leu-Arg-Phe-amide) in arthropods (Tanaka, 2016) and of pQDVDHVFLRFamide (pQ-Asp-Val-Asp-His-Val-Phe-Leu-Arg-Phe-amide) in insects (Veenstra et al., 2012) (X = any amino acid; pQ = pyroglutamate). It is believed that the myosuppressin and FMRFamide are evolutionarily related (Veenstra et al., 2012). Similarly, two

genes were found that resemble these genes from insects. Also, one putative gene for myosuppressin receptor was found. The receptor sequence was discovered using the query of myosuppressin receptor for *D. melanogaster*, and the resultant gene sequence had a blast e-value and percent identity of 2e-44 and 50.87 % for both *V. destructor* and *V. jacobsoni,* respectively. The sequence was discovered by back blasting against *D. melanogaster* genomic database and phylogenetic clustering with 90% support in neighborhood joining tree method (1000 bootstrapping) with known myosuppressin receptor sequences from closely related arthropods.

Allatostatin B receptor:

The neuropeptide allatostatin B neuropeptide, also known as myoinhibitory peptide, was first discovered in crickets, *Grylliodea* (Williamson et al., 2001). By regulating the Corpora allata (CA), Ast-B signaling functions for inhibition of Juvenile hormone. Two gene sequences encoding the Ast-B receptor were identified in the each of the *Varroa* mite species genomic scaffold, with blast e-value between 3e-95 and 5e-112, and percent identity between 40 % - 45 %, when blast searched using Ast-B receptor of *D. melanogaster*. Final confirmation of the sequences was done using phylogenetic clustering with 52% bootstrapping support in neighborhood joining tree method (1000 bootstrapping) with known Ast-B receptor sequences from closely related arthropods. Seven TMs were located in the receptor sequence (Figure 3.19A), indicating that the gene prediction was accurate.

Trissin receptor:

Trissin is a cysteine rich, non-amidated, 27 amino acids long neuropeptide with an intramolecular disulfide bridge that typically contains the CCFYNL amino acid motif (Cys-Phe-Tyr-Asn-Leu) at the C-terminus of the mature peptide in various arthropods (Ida et al., 2011b). Trissin is newly identified group of neuropeptides in insects. It was first identified from *D.*

melanogaster (Ida et al., 2011b) and subsequently in the silk moth, *B. mori,* (Roller et al., 2016) as well as in other arthropods such as crayfish, *Procambarus clarkii,* and spider mite, *T. urticae* (Li et al., 2014; Veenstra et al., 2012). Although, the location of the peptide's expression is known, its biological function in various arthropods is still unknown. One gene sequence potentially encodes for the putative Trissin receptor was identified from the genomes of *V. destructor* and *V. jacobsoni*. The identified gene sequence shares moderate sequence homology with the Trissin receptor sequences from *D. melanogaster,* with a blast e-value of 1e-99 and a percent identity of 38%. The sequence was confirmed by back blasting against genomic database of *D. melanogaster* and followed by phylogenetic clustering with 51% bootstrapping support in neighborhood joining tree method (1000 bootstrapping) with the known receptor sequences from closely related arthropods.

EFLamide:

EFLamide (Glu-Phe-Leu-amide) is an arthropod neuropeptide ortholog of vertebrate thyrotropic releasing hormone (TRH). The EFLamide gene is thought to encode two different type of peptide, namely EFLamide and EFLGGPamide (Glu-Phe-Leu-Gly-Gly-Pro-amide), in various species (Veenstra et al., 2012). Among arthropods, EFLamide gene was initially identified in the spider mite, *T. urticae,* and was later discovered in other chelicerates such as ticks, *I. scapularis* and *Rhipicephalus spp.,* the mite *Sarcoptes scabiei* , the scorpion, *Hadrurus gertschi*, crustaceans such as *Daphnia pulex* and *Marsupeneus japonicus* (Veenstra et al., 2012) and insects such as *Locusta migratoria, Cimex lectularis, Anax spp* (Veenstra, 2019b; Veenstra and Šimo, 2020). By using the *L. migratoria*'s EFLamide receptor sequence as a query in the blast search, two gene sequences encoding the cognate receptor were identified in the genomic database of *V. destructor* and *V. jacobsoni.* The sequence found were confirmed by back blasting

and phylogenetic clustering with 50% bootstrapping support in neighborhood joining tree method (1000 bootstrapping) with known EFLamide receptor sequences from closely related arthropods.

Class B GPCRs

Diuretic Hormone type peptide receptors

This peptide group includes the calcitonin-like/diuretic hormone 31 type and Corticotropin releasing factor-like/diuretic hormone 44 type. The diuretic hormone 31 type is a neuropeptide with a 31 amino acids long residue and has limited similarity with the vertebrate calcitonin family (Veenstra et al., 2012). In insects, it was first observed in cockroach, *Diploptera punctate* (Coast and Schooley, 2011; Furuya et al., 2000). The ortholog neuropeptide and receptor gene sequence were easily identified from the genomic database of both *Varroa* mites. For each mite species, the resultant receptor gene sequence had a blast e-value of 2e-89 and a percent identity of a 47%. The receptor sequence (Figure 3.22A) was analyzed for the presence and location of seven TMs.

Diuretic hormone 44 has a 44 long amino acid residue and, unlike DH 31 type, is similar to vertebrate corticotropin-releasing factor-like. These neuropeptides are believed to act as a diuretic by stimulating Malpighian tubule secretion. A single gene sequence that encodes for the cognate receptor of DH 44 type was identified in both *V. destructor* and *V. jacobsoni*. The identified sequences had a blast e-values of 1e-96 and 2e-100, respectively. Clustering (neighborhood joining method with 1000 bootstrapping tests) the potential DH 44 receptor gene sequence with orthologous DH 44 receptor gene sequences from closely related arthropods helped in confirming the results. The sequence (Figure 3.23A) was analyzed for the presence and location of seven TMs.

Pigment dispersing factor:

The pigment dispersing factor (PDF) closely related with the crustacean pigment dispersing hormone, but is associated with a different role in arthropods in regulating activities related to the circadian rhythmic clock like feeding, etc. (Renn et al., 1999). Due to the low level of conservation of the PDF sequence (Dircksen et al., 2011), finding the gene sequence using the blast search is difficult. By searching in the SRA database of each *Varroa* mite species, gene sequence encoding the PDF neuropeptide gene sequence was found. Also, one receptor gene sequence was identified that had orthologous relationships with the PDF receptor of *D. melanogaster*. The identified sequence had a blast e-value of 3e-50 in both mites. Results were confirmed by phylogenetic clustering with 67% bootstrapping support in neighborhood joining method (1000 bootstrapping), with known PDF receptor sequences from closely related arthropods. The receptor sequence (Figure 3.24A) was analyzed for the presence and location of seven TMs.

Insect Parathyroid Hormone:

Parathyroid hormone (PTH), a vertebrate neuropeptide that is highly expressed in kidneys and bones, is involved in calcium regulation, phosphate hemostasis, and bone growth (Abou-Samra et al., 1994; Li et al., 2013). Among species other than mammals, PTH signaling was first found in red flour moth, *T. castaneum* (Li et al., 2013). No sequence was found that encodes PTH neuropeptide gene from the genomic database of *Varroa* mite. However, three sequences encoding PTH receptor orthologies were identified, which is in parallel with a recent report of PTH signaling *T. castaneum* (Xie et al., 2020). The genes found in *Varroa* mites have a sequence similarity of 37.85%, 37.53 and 36.54% (*V. destructor*) and 36.92%, 37.59%, and 36.54% (*V. jacobsoni*) to PTHR1 of *T. castaneum* (XP_015834182.1). Final confirmation was done by

clustering the gene sequences with 60% bootstrapping support in neighborhood joining tree method (1000 bootstrapping), with know iPTH receptor sequences from the closely related arthropods. Seven TMs were located in the receptor sequence (Figure 3.25A).

GPCRs not found

We could not find the orthologies for following GPCRs: CNMamide (Cys-Asn-Met-amide) receptor, Elevenin receptor, ETH receptor, AKH receptor, Arginine vasopressin like peptide receptor. However, the neuropeptide sequence of the CNMamide and AVPL in each of *Varroa* mite species were identified. Additionally, the following receptor gene sequence were found but not their cognate neuropeptide gene sequence: pyrokinin receptor, gonadulin receptor, myosuppressin (*V. destructor*), and sNPF receptor. These anomalies may indicate that errors occurred during genome assembly or that a portion of the genome assembly is missing.

Gonadotropin-releasing hormone (GnRH) is a vertebrate hormone, usually involved in reproduction control via regulation of gonadotropins luteinizing hormone (LH) and follicle-stimulating hormone (Kah et al., 2007). As the mechanisms of reproduction are similar across the animal kingdom, it is believed that the regulation of reproduction is highly conserved in the ancestral lineage of the protostomes and deuterostomes. GnRH like peptides have also been identified in the protostomian invertebrates belonging to the Phylum Lophotrochozoa, including *Sepia officinalis, Aplysia californica, and Lottia gigantea* (mollusks) (Di Cristo et al., 2009; Tsai and Zhang, 2008; Zhang et al., 2000) and *Capitella teleta*, and *Helobdella robusta* (annelids) (Roch et al., 2011; Tsai and Zhang, 2008). In contrast to lophotrochozoans, ecdysozoans employ ligands different from the GnRH peptides, for binding with retained GnRHR-like receptors (Tsai and Zhang, 2008). AKH was initially identified as a possible ligand for GnRH-like receptors in *D. melanogaster* and *B. morri.* AKH's role as a ligand for GnRHR-like receptors was proved

with the finding of AKH-GnRH-like neuropeptide in nematode *Caenorhabditis elegans* that activates the GnRHR-like receptor, *Ce*GnRHR, in a dose dependent manner. The neuropeptide found shared similarities with GnRH-like ligands found in higher order vertebrates and AKH of insects (Lindemans et al., 2011).

Corazonin is another peptide that is structurally similar to AKH and is believed to be another potential ligands of GnRHR-like receptors in ecdysozoans. The phylogenetic relationships between these neuropeptides and their cognate receptors indicate that they are closely related and evolved as a result of the neuropeptide-receptor coevolution via ancestral gene duplication for both neuropeptide and receptor. Although AKH and corazonin are related, they do not exhibit cross reactivity with their respective receptor, implying the existence of two independent hormonal systems. This property aided in the identification of a third closely related neuropeptide, ACP, which was initially discovered in *A. gambiae* (Hansen et al., 2010). ACP and its receptor were discovered to be a structural intermediates between the AKH and corazonin. Although ACP and receptor are found in a wide variety of insects, they are absent from the reference species *D. melanogaster* and *A. mellifera.*

In one reference species, *I. scapularis*, the AKH neuropeptide and receptor gene are absent, whereas three GnRH-like neuropeptides and five AKH and ACP-like receptors were discovered in another reference species, *T. urticae* (Veenstra et al., 2012). Although ACP and corazonin signaling were detected in *Varroa* mite, AKH signaling was missing. It is difficult to determine whether the *Varroa* mite actually lacks AKH signaling, as observed in the closely related *T. urticae*. Also, the presence of a receptor and the absence of a neuropeptide raises our suspicions about the quality of *Varroa* mite genome assembly. According to the theory of ligand-receptor coevolution theory, it is very unlikely that a neuropeptide gene is absent while its

receptor is present, and vice versa. Thus, further confirmation is required to determine whether AKH signaling is indeed absent from the *Varroa mite* genome assembly. Likewise, more confirmation is required about the absence of neuropeptide receptors including Elevenin, ETH, CNMamide, and AVPL.

References:

1. Abou-Samra, A., Jüppner, H., Kong, X., Schipani, E., Iida-Klein, A., Karga, H., Urena, P., Gardella, T., Potts, J.J., Kronenberg, H., 1994. Structure, function, and expression of the receptor for parathyroid hormone and parathyroid hormone-related peptide. Advances in nephrology from the Necker Hospital 23, 247-264.
2. Armenteros, J.J.A., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature biotechnology 37, 420-423.
3. Ayub, M., Hermiz, M., Lange, A.B., Orchard, I., 2020. SIFamide influences feeding in the Chagas disease vector, Rhodnius prolixus. Frontiers in neuroscience 14, 134.
4. Birgül, N., Weise, C., Kreienkamp, H.J., Richter, D., 1999. Reverse physiology in Drosophila: identification of a novel allatostatin-like neuropeptide and its cognate receptor structurally related to the mammalian somatostatin/galanin/opioid receptor family. The EMBO Journal 18, 5892-5900.
5. Caers, J., Verlinden, H., Zels, S., Vandersmissen, H.P., Vuerinckx, K., Schoofs, L., 2012. More than two decades of research on insect neuropeptide GPCRs: an overview. Frontiers in endocrinology 3, 151.
6. Collin, C., Hauser, F., Krogh-Meyer, P., Hansen, K.K., de Valdivia, E.G., Williamson, M., Grimmelikhuijzen, C.J., 2011. Identification of the Drosophila and Tribolium receptors for the recently discovered insect RYamide neuropeptides. Biochemical and biophysical research communications 412, 578-583.
7. Cazzamali, G., Hauser, F., Kobberup, S., Williamson, M., Grimmelikhuijzen, C.J., 2003. Molecular identification of a Drosophila G protein-coupled receptor specific for crustacean cardioactive peptide. Biochemical and biophysical research communications 303, 146-152.
8. Christie, A.E., 2008. Neuropeptide discovery in Ixodoidea: an in silico investigation using publicly accessible expressed sequence tags. General and comparative endocrinology 157, 174-185.
9. Christie, A.E., 2015. In silico characterization of the neuropeptidome of the Western black widow spider Latrodectus hesperus. General and comparative endocrinology 210, 63-80.
10. Coast, G.M., Schooley, D.A., 2011. Toward a consensus nomenclature for insect neuropeptides and peptide hormones. Peptides 32, 620-631.
11. Crooks, G.E., Hon, G., Chandonia, J.-M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome research 14, 1188-1190.
12. Dircksen, H., Neupert, S., Predel, R., Verleyen, P., Huybrechts, J., Strauss, J., Hauser, F., Stafflinger, E., Schneider, M., Pauwels, K., 2011. Genomics, transcriptomics, and peptidomics of Daphnia pulex neuropeptides and protein hormones. Journal of proteome research 10, 4478-4504.

13. Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32, 1792-1797.
14. Furuya, K., Milchak, R.J., Schegg, K.M., Zhang, J., Tobe, S.S., Coast, G.M., Schooley, D.A., 2000. Cockroach diuretic hormones: characterization of a calcitonin-like peptide in insects. Proceedings of the National Academy of Sciences 97, 6469-6474.
15. Garczynski, S.F., Brown, M.R., Shen, P., Murray, T.F., Crim, J.W., 2002. Characterization of a functional neuropeptide F receptor from Drosophila melanogaster. Peptides 23, 773-780.
16. Groome, J.R., Tillinghast, E.K., Townley, M.A., Vetrovs, A., Watson III, W.H., Hunt, D.F., Griffin, P.R., Alexander, J.E., Shabanowitz, J., 1990. Identification of proctolin in the central nervous system of the horseshoe crab, Limulus polyphemus. Peptides 11, 205-211.
17. Hansen, K.K., Hauser, F., Williamson, M., Weber, S.B., Grimmelikhuijzen, C.J., 2011. The Drosophila genes CG14593 and CG30106 code for G-protein-coupled receptors specifically activated by the neuropeptides CCHamide-1 and CCHamide-2. Biochemical and biophysical research communications 404, 184-189.
18. Hansen, K.K., Stafflinger, E., Schneider, M., Hauser, F., Cazzamali, G., Williamson, M., Kollmann, M., Schachtner, J., Grimmelikhuijzen, C.J., 2010. Discovery of a novel insect neuropeptide signaling system closely related to the insect adipokinetic hormone and corazonin hormonal systems. Journal of Biological Chemistry 285, 10736-10747.
19. Hauser, F., Nothacker, H.-P., Grimmelikhuijzen, C.J., 1997. Molecular cloning, genomic organization, and developmental regulation of a novel receptor from Drosophila melanogaster structurally related to members of the thyroid-stimulating hormone, follicle-stimulating hormone, luteinizing hormone/choriogonadotropin receptor family from mammals. Journal of Biological Chemistry 272, 1002-1010.
20. Hauser, F., Cazzamali, G., Williamson, M., Blenau, W., Grimmelikhuijzen, C.J., 2006. A review of neurohormone GPCRs present in the fruitfly Drosophila melanogaster and the honey bee Apis mellifera. Progress in neurobiology 80, 1-19.
21. Jékely, G., 2013. Global view of the evolution and diversity of metazoan neuropeptide signaling. Proceedings of the National Academy of Sciences 110, 8702-8707.
22. Jiang, H., Kim, D., Dobesh, S., Evans, J.D., Nachman, R.J., Kaczmarek, K., Zabrocki, J., Park, Y., 2016. Ligand selectivity in tachykinin and natalisin neuropeptidergic systems of the honey bee parasitic mite *VarroaVarroa destructor*. Scientific reports 6, 19547.
23. Jurenka, R., 2015. The PRXamide neuropeptide signalling system: conserved in animals, Advances in Insect Physiology. Elsevier, pp. 123-170.
24. Kramer, S.J., ToSCHI, A., Miller, C.A., Kataoka, H., Quistad, G.B., Li, J.P., Carney, R.L., Schooley, D.A., 1991. Identification of an allatostatin from the tobacco hornworm Manduca sexta. Proceedings of the national academy of sciences 88, 9458-9462.
25. Lee, D., Broeck, J.V., Lange, A.B., 2013. Identification and expression of the CCAP receptor in the Chagas' disease vector, Rhodnius prolixus, and its involvement in cardiac control. PloS one 8, e68897.
26. Li, B., Predel, R., Neupert, S., Hauser, F., Tanaka, Y., Cazzamali, G., Williamson, M., Arakane, Y., Verleyen, P., Schoofs, L., 2008. Genomics, transcriptomics, and peptidomics of neuropeptides and protein hormones in the red flour beetle Tribolium castaneum. Genome research 18, 113-122.
27. Li, C., Chen, M., Sang, M., Liu, X., Wu, W., Li, B., 2013. Comparative genomic analysis and evolution of family-B G protein-coupled receptors from six model insect species. Gene 519, 1-12.

28. Li, L., Kelley, W.P., Billimoria, C.P., Christie, A.E., Pulver, S.R., Sweedler, J.V., Marder, E., 2003. Mass spectrometric investigation of the neuropeptide complement and release in the pericardial organs of the crab, Cancer borealis. Journal of neurochemistry 87, 642-656.
29. Li, S., Hauser, F., Skadborg, S.K., Nielsen, S.V., Kirketerp-Møller, N., Grimmelikhuijzen, C.J., 2016. Adipokinetic hormones and their G protein-coupled receptors emerged in Lophotrochozoa. Scientific reports 6, 1-13.
30. Loi, P., Emmal, S., Park, Y., Tublitz, N., 2001. Identification, sequence and expression of a crustacean cardioactive peptide (CCAP) gene in the moth Manduca sexta. Journal of Experimental Biology 204, 2803-2816.
31. Maule, A., Shaw, C., Halton, D., Thim, L., Johnston, C., Fairweather, I., Buchanan, K., 1991. Neuropeptide F: a novel parasitic flatworm regulatory peptide from Moniezia expansa (Cestoda: Cyclophyllidea). Parasitology 102, 309-316.
32. Nagata, S., 2016. FMRFamides, Handbook of Hormones. Elsevier, pp. 432-433.
33. Nässel, D.R., Wegener, C., 2011. A comparative review of short and long neuropeptide F signaling in invertebrates: Any similarities to vertebrate neuropeptide Y signaling? Peptides 32, 1335-1355.
34. Nässel, D.R., 2018. Substrates for neuronal cotransmission with neuropeptides and small molecule neurotransmitters in Drosophila. Frontiers in cellular neuroscience 12, 83.
35. Nässel, D.R., Zandawala, M., 2019. Recent advances in neuropeptide signaling in Drosophila, from genes to physiology and behavior. Progress in neurobiology 179, 101607.
36. Nguyen, T.V., Rotllant, G.E., Cummins, S.F., Elizur, A., Ventura, T., 2018. Insights into sexual maturation and reproduction in the Norway lobster (Nephrops norvegicus) via in silico prediction and characterization of neuropeptides and G protein-coupled receptors. Frontiers in endocrinology 9, 430.
37. Orchard, I., Lee, D.H., Da Silva, R., Lange, A.B., 2011. The proctolin gene and biological effects of proctolin in the blood-feeding bug, Rhodnius prolixus. Frontiers in endocrinology 2, 59.
38. Nusbaum, M.P., Blitz, D.M., Marder, E., 2017. Functional consequences of neuropeptide and small-molecule co-transmission. Nature Reviews Neuroscience 18, 389.
39. Park, Y., Kim, Y.-J., Adams, M.E., 2002. Identification of G protein-coupled receptors for Drosophila PRXamide peptides, CCAP, corazonin, and AKH supports a theory of ligand-receptor coevolution. Proceedings of the National Academy of Sciences 99, 11423-11428.
40. Renn, S.C., Park, J.H., Rosbash, M., Hall, J.C., Taghert, P.H., 1999. A pdf neuropeptide gene mutation and ablation of PDF neurons each cause severe abnormalities of behavioral circadian rhythms in Drosophila. Cell 99, 791-802.
41. Roller, L., Yamanaka, N., Watanabe, K., Daubnerová, I., Žitňan, D., Kataoka, H., Tanaka, Y., 2008. The unique evolution of neuropeptide genes in the silkworm Bombyx mori. Insect biochemistry and molecular biology 38, 1147-1157.
42. Sellami, A., Veenstra, J.A., 2015. SIFamide acts on fruitless neurons to modulate sexual behavior in Drosophila melanogaster. Peptides 74, 50-56.
43. Šimo, L., Itňan, D.Ž., Park, Y., 2009. Two novel neuropeptides in innervation of the salivary glands of the black-legged tick, Ixodes scapularis: Myoinhibitory peptide and SIFamide. Journal of Comparative Neurology 517, 551-563.
44. Šimo, L., Koči, J., Park, Y., 2013. Receptors for the neuropeptides, myoinhibitory peptide and SIFamide, in control of the salivary glands of the blacklegged tick Ixodes scapularis. Insect biochemistry and molecular biology 43, 376-387.

45. Sonnhammer, E.L., Von Heijne, G., Krogh, A., 1998. A hidden Markov model for predicting transmembrane helices in protein sequences, Ismb, pp. 175-182.
46. Spittaels, K., Vankeerberghen, A., Torrekens, S., Devreese, B., Grauwels, L., Van Leuven, F., Hunt, D., Shabanowitz, J., Schoofs, L., Van Beeumen, J., 1995. Isolation of Ala1-proctolin, the first natural analogue of proctolin, from the brain of the Colorado potato beetle. Molecular and cellular endocrinology 110, 119-124.
47. Stangier, J., Hilbich, C., Beyreuther, K., Keller, R., 1987. Unusual cardioactive peptide (CCAP) from pericardial organs of the shore crab Carcinus maenas. Proceedings of the National Academy of Sciences 84, 575-579.
48. Starratt, A.N., Brown, B.E., 1975. Structure of the pentapeptide proctolin, a proposed neurotransmitter in insects. Life sciences 17, 1253-1256.
49. Thiel, D., Bauknecht, P., Jékely, G., Hejnol, A., 2019. A nemertean excitatory peptide/CCHamide regulates ciliary swimming in the larvae of Lineus longissimus. Frontiers in zoology 16, 28.
50. Tobe, S.S., Bendena, W.G., 2013. Allatostatins, Handbook of Biologically Active Peptides. Elsevier, pp. 191-196.
51. Toullec, J.-Y., Corre, E., Mandon, P., Gonzalez-Aravena, M., Ollivaux, C., Lee, C.-Y., 2017. Characterization of the neuropeptidome of a Southern Ocean decapod, the Antarctic shrimp Chorismus antarcticus: Focusing on a new decapod ITP-like peptide belonging to the CHH peptide family. General and Comparative Endocrinology 252, 60-78.
52. Xie, J., Sang, M., Song, X., Zhang, S., Kim, D., Veenstra, J.A., Park, Y., Li, B., 2020. A new neuropeptide insect parathyroid hormone iPTH in the red flour beetle Tribolium castaneum. PLoS Genetics 16, e1008772.
53. Veenstra, J.A., 2000. Mono-and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors. Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America 43, 49-63.
54. Veenstra, J.A., 2009a. Allatostatin C and its paralog allatostatin double C: the arthropod somatostatins. Insect biochemistry and molecular biology 39, 161-170.
55. Veenstra, J.A., 2009b. Does corazonin signal nutritional stress in insects? Insect biochemistry and molecular biology 39, 755-762.
56. Veenstra, J.A., 2016. Neuropeptide evolution: Chelicerate neurohormone and neuropeptide genes may reflect one or more whole genome duplications. General and comparative endocrinology 229, 41-55.
57. Veenstra, J.A., Rombauts, S., Grbić, M., 2012. In silico cloning of genes encoding neuropeptides, neurohormones and their putative G-protein coupled receptors in a spider mite. Insect biochemistry and molecular biology 42, 277-295.
58. Zels, S., Verlinden, H., Dillen, S., Vleugels, R., Nachman, R.J., Broeck, J.V., 2014. Signaling properties and pharmacological analysis of two sulfakinin receptors from the red flour beetle, Tribolium castaneum. PLoS One 9, e94502.
59. Williamson, M., Lenz, C., Winther, M., Nässel, D.R., Grimmelikhuijzen, C.J., 2001. Molecular cloning, genomic organization, and expression of a B-type (cricket-type) allatostatin preprohormone from Drosophila melanogaster. Biochemical and biophysical research communications 281, 544-550.

Figures

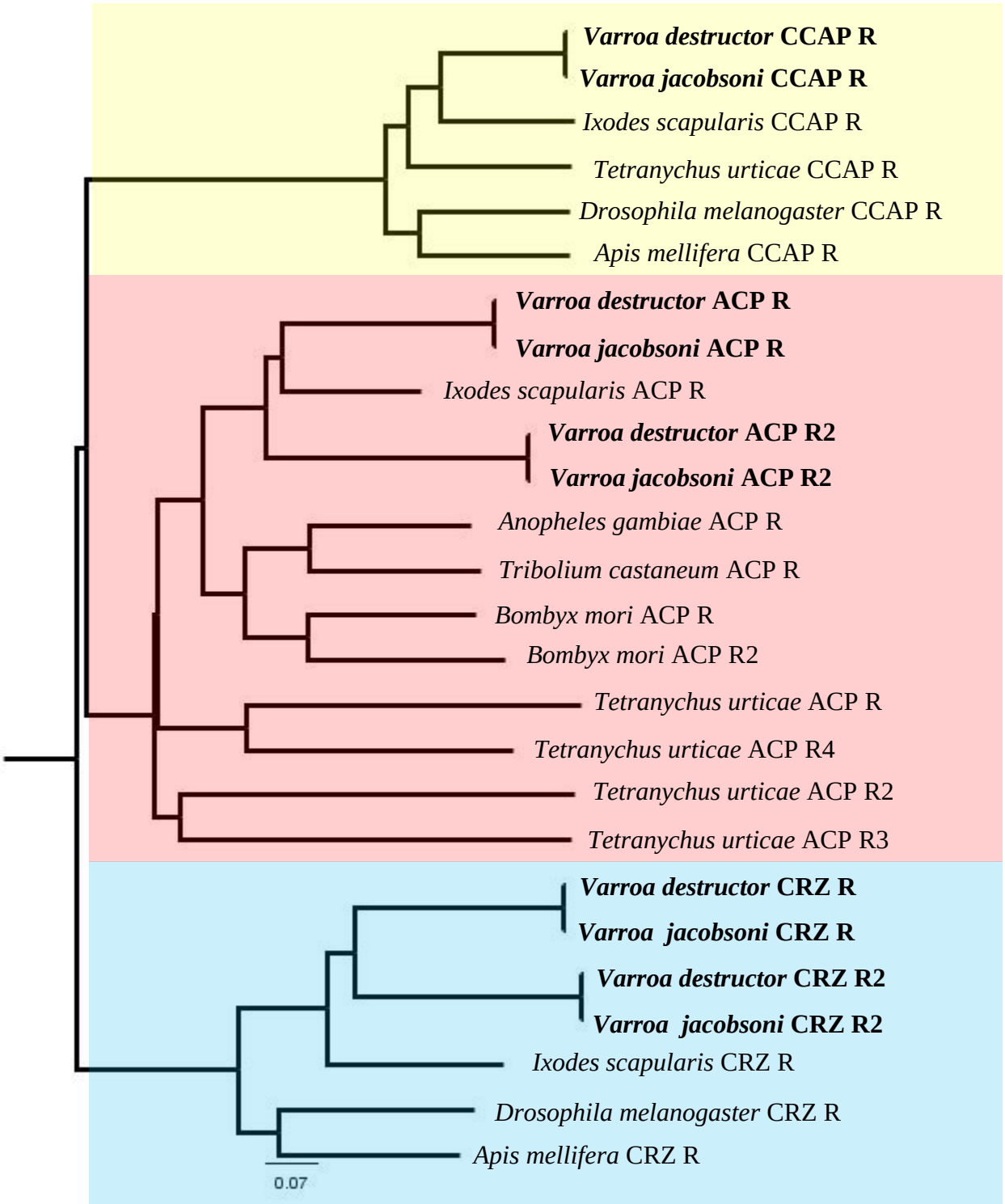

***Varroa destructor* CCAP R**
***Varroa jacobsoni* CCAP R**
Ixodes scapularis CCAP R
Tetranychus urticae CCAP R
Drosophila melanogaster CCAP R
Apis mellifera CCAP R
***Varroa destructor* ACP R**
***Varroa jacobsoni* ACP R**
Ixodes scapularis ACP R
***Varroa destructor* ACP R2**
***Varroa jacobsoni* ACP R2**
Anopheles gambiae ACP R
Tribolium castaneum ACP R
Bombyx mori ACP R
Bombyx mori ACP R2
Tetranychus urticae ACP R
Tetranychus urticae ACP R4
Tetranychus urticae ACP R2
Tetranychus urticae ACP R3
***Varroa destructor* CRZ R**
***Varroa jacobsoni* CRZ R**
***Varroa destructor* CRZ R2**
***Varroa jacobsoni* CRZ R2**
Ixodes scapularis CRZ R
Drosophila melanogaster CRZ R
Apis mellifera CRZ R
0.07

Figure 3.2 A. Phylogenetic tree having ACP receptor, Corazonin receptor, CCAP receptor clusters made species from Chelicerata, Crustacea and Hexapoda Class. The receptors for *V. destructor* and *V. jacobsoni*. are highlighted with bold font. The clades for different receptors are marked with different colors. The tree was constructed using neighbor-joining tree using 1000 bootstrapping and poisson substitution method.

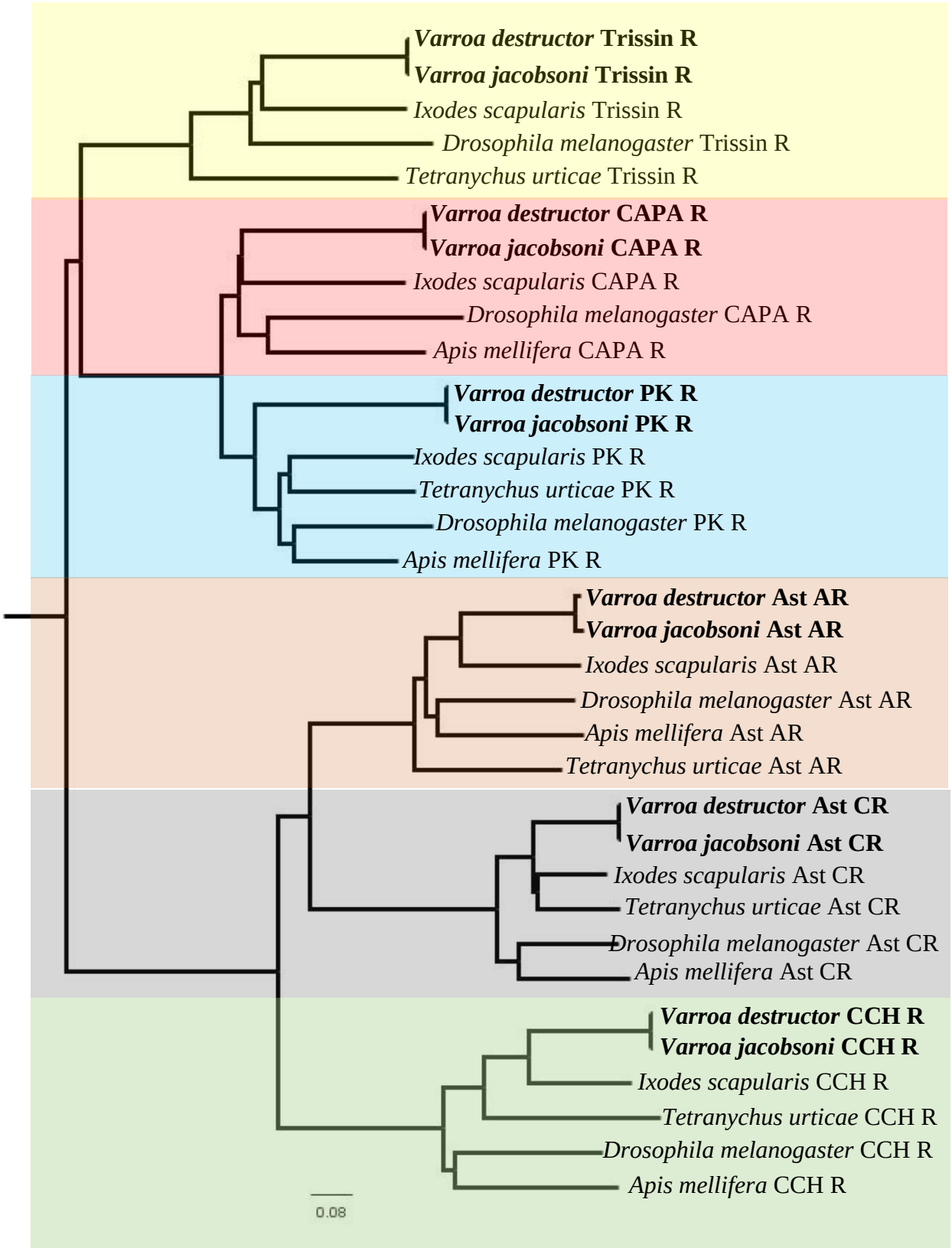


Figure 3.3 A. Phylogenetic tree having Ast A and C receptor, CAPA receptor, CCH receptor, Pyrokinin receptor and Trissin receptor clusters made species from Chelicerata,

Crustacea and Hexapoda Class. The receptors for *V. destructor* and *V. jacobsoni*. are highlighted with bold font. The clades for different receptors are marked with different colors. The tree was constructed using neighbor-joining tree using 1000 bootstrapping and poisson substitution method.

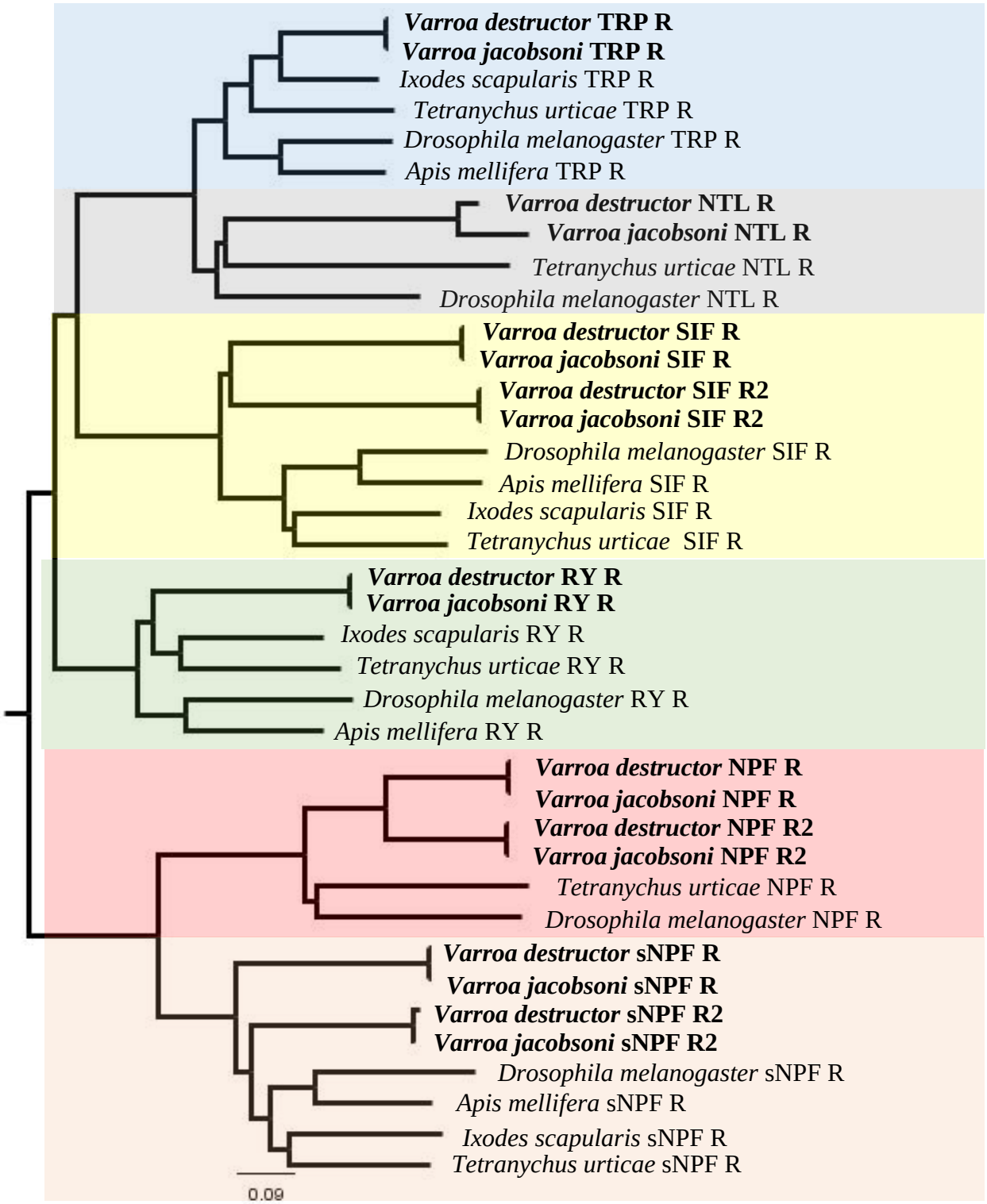


Figure 3.4 Phylogenetic tree having Tachykin (TRP) receptor, Natalisin receptor, NPF receptor, sNPF receptor, SIFamide receptor and RYamide receptor clusters made species from Chelicerata, Crustacea and Hexapoda Class. The receptors for *V. destructor* and *V. jacobsoni* are highlighted in bold font. The tree was constructed using neighbor-joining tree using 1000 bootstrapping and poisson substitution method.

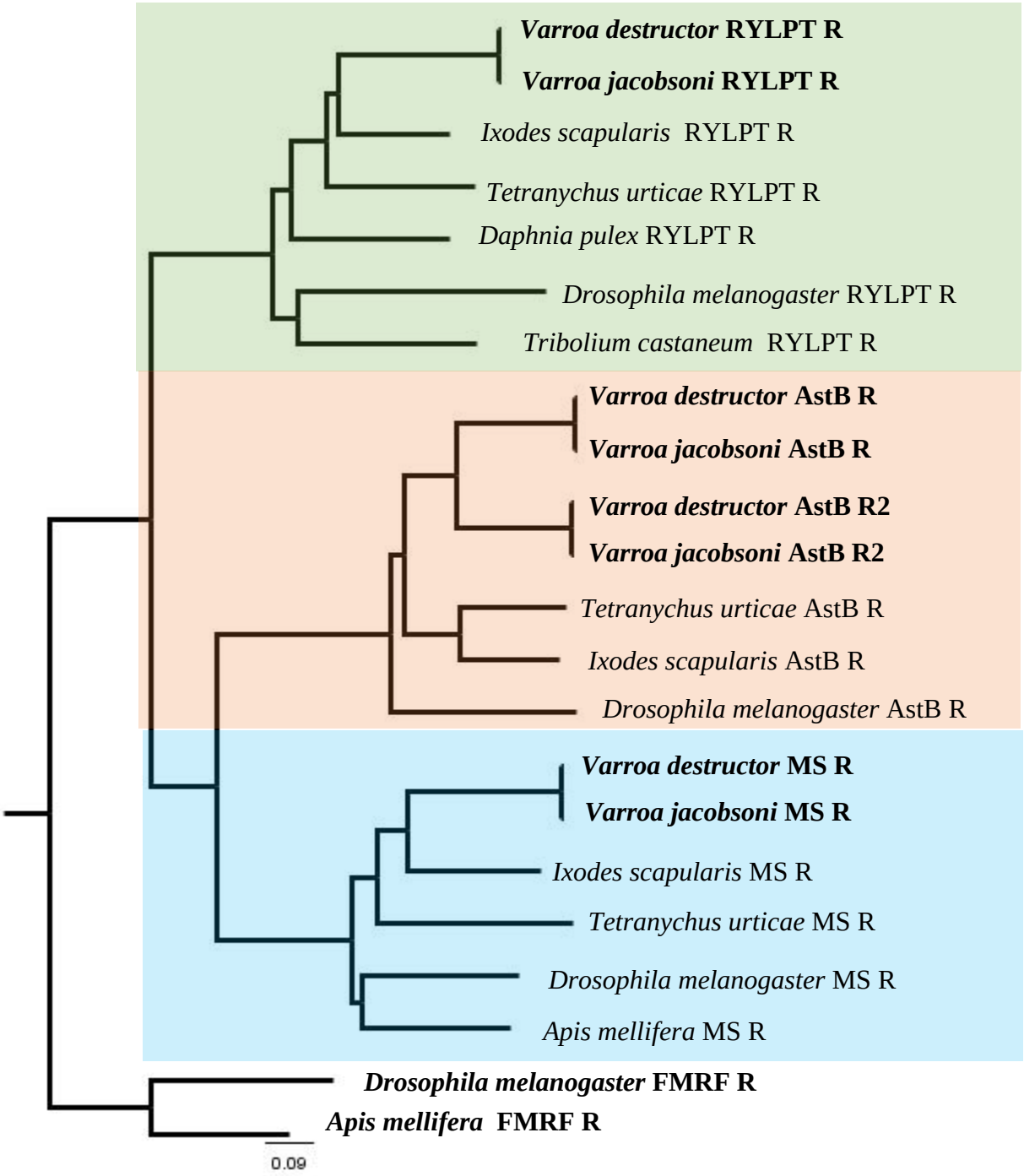


Figure 3.5 Phylogenetic tree having FMRFamide receptors, Myosuppressin receptor, Allatostatin B and Proctolin receptor cluster. Different receptor clusters are marked with different colors. Various receptors for *V. destructor* and *V. jacobsoni* are highlighted with bold font. The tree was constructed using neighbor-joining tree using 1000 bootstrapping and poisson substitution method.

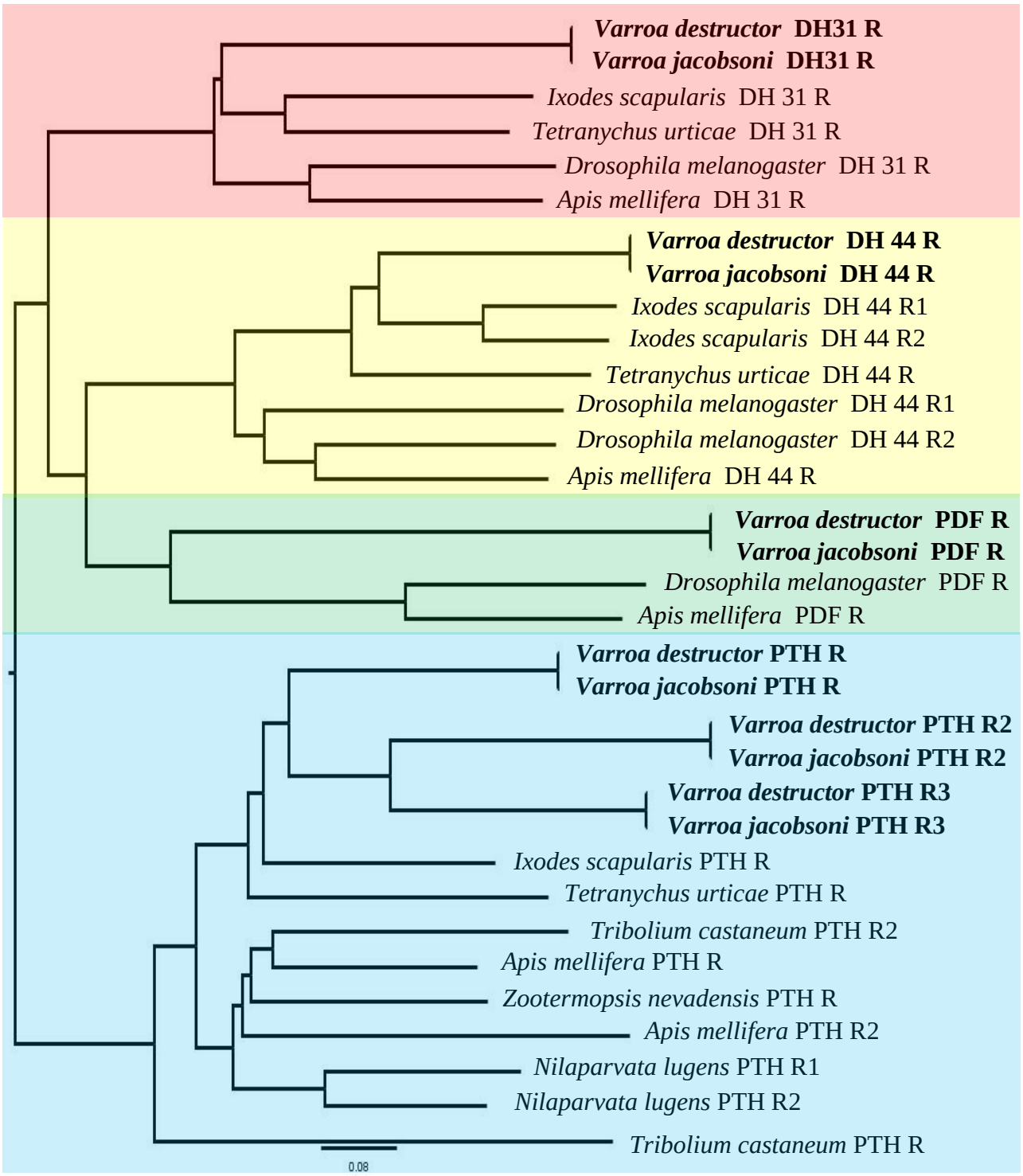


Figure 3.6 Phylogenetic tree having Diuretic hormone 31R, Diuretic hormone 44R, Pigment dispersing factor receptor, and Parathyroid hormone receptor. Different receptor clusters are marked with different colors. DH 31R, DH 44R, PTHR, and PDFR from mite species are highlighted with Bold font. The tree was constructed using neighbor-joining tree using 1000 bootstrapping and poisson substitution method.

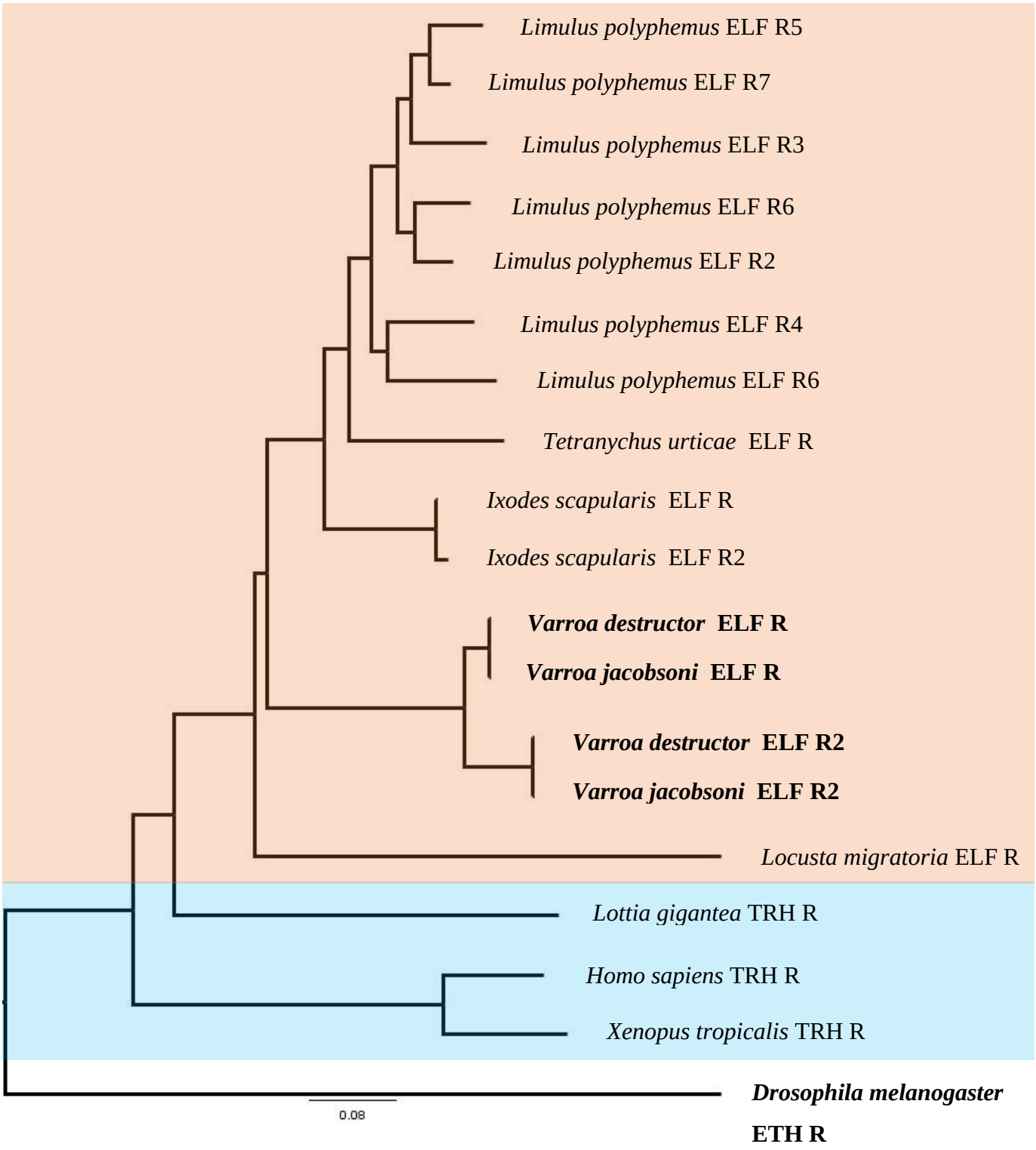


Figure 3.7 Phylogenetic tree having EFLamide receptors, TRH receptor and ETH receptor. Different receptor clusters are marked with different colors. EFLamide receptor from mite species are highlighted with Bold font. The tree was constructed using neighbor-joining tree using 1000 bootstrapping and poisson substitution method.

Table 3.1 Master table for neuropeptide receptor sequences.

	Varroa destructor		*Varroa jacobsoni*					
Neuropeptide receptor	**Accession number**	**Ligand Gene ID**	**Accession number**	**Ligand Gene ID**	***D. melanogaster***	***A. mellifera***	***Ixodes scapularis***	***Tetranychus urticae***
ACP	XP_022652899.1 XP_022655491.1	LOC111255317	XP_022699658.1 XP_022697283.1	LOC111255317	ND	ND	+	ND
AKH	ND*	ND*	ND*	ND*	+	+	ND	+
Ast A	XP_022649205.1	LOC111245583	XP_022689205.1	LOC111266071	+	+	+	+
Ast B/MIP	XP_022666601.1 XP_022668353.1	LOC111251198	XP_022700716.1 XP_022693585.1	LOC111271347	+	+	+	+
Ast C	XP_022653392.1	SRR8864012.160068737.1	XP_022700513.1	LOC111262149	+	+	+	+
AVPL	ND*	LOC111251516	ND*	LOC111260306	ND	ND	+	+
Burs/rk	XP_022651891.1	LOC111247436, LOC111264752	XP_022702359.1	LOC111264753, LOC111264752	+	+	+	+
DH31	XP_022663535.1	LOC111249899	XP_022697774.1	SRR3632582.162967731.2	+	+	+	+
CAPA	XP_022668622.1	LOC111250488	XP_022698878.1	LOC111259753	+	+	+	ND*
CCH amide	XP_022651431.1	LOC111251153	XP_022693677.1	LOC111270805	+	+	+	+
CNMamide	ND*	SRR8864012.46893032.2	ND*	SRR3632582.185051089.2	+	ND*	ND*	ND*

Corazonin	XP_022665765.1 XP_022668003.1 XP_022655704.1	LOC 11125154	XP_022687694.1 XP_022690575.1 XP_022701755.1	SRR8864012.127467345.2	+	+	+	ND*
CRF like-DH (DH37/47)	ND*	ND*	ND*	ND*	+	+	+	+
CCAP	XP_022663053.1	LOC111251821	XP_022697696.1	LOC111263999	+	+	+	+
DH 44	XP_022652862.1	LOC111248063	XP_022693806.1	LOC111262624	+	+	+	+
EFLamide	XP_022671024.1 XP_022671503.1	ND*	XP_022705342.1 XP_022707492.1	ND*	ND*	ND*	+	+
Elevenin	ND*	SRR8864012.149089260.1	ND*	ND*	+	+	+	+
ETH	ND*	ND*	ND*	ND*	+	+	+	+
FMRFamide	ND*	LOC111249730	ND*	LOC111260277	+	+	+	ND*
Gonadulin	XP_022666433.1	ND*	XP_022685875.1	ND*	+	+	+	+
GPH	ND*	ND*	ND*	SRR8100122.28924566.2 SRR3632582.77147140.2	+	ND*	+	+
Kinin	ND*	ND*	ND*	ND*	+	ND*	+	+
MS	XP_022652803.1	ND*	XP_022701117.1	SRR3635105.57299974.2	+	+	+	+
Natalisin	AKR03971.1	LOC111251926	XP_022705660.1	LOC111262523	+	ND*	+	+
NPLP	XP_022665141.1	ND*	XP_022700676.1	ND*	+	+	+	+
NPF	XP_022671156.1 XP_022671227.1	SRR5377265.26255811.2	XP_022689861.1 XP_022702941.1	SRR3632582.185285521.2	+	ND*	ND*	+
Orcokinin	ND^	ND*	ND^	SRR3632582.136591231.1	ND^	ND^	ND^	ND^

PDF	XP_022654705.1	SRR8864012.106407301.1	XP_022702049.1	SRR11879882.65162670.2	+	+	ND*	ND*
Proctolin	XP_022663222.1	LOC111244398	XP_022707176.1	SRR3635105.58662445.1	+	ND*	+	+
iPTH	XP_022653897.1 XP_022663980.1 XP_022663823.1	ND*	XP_022711005.1 XP_022707175.1 XP_022699433.1	ND*	ND*	+	+	+
PK	XP_022654818.1	ND*	XP_022704826.1	ND*	+	+	+	+
RYamide	XP_022643695.1	ND*	XP_022694295.1	ND*	+	+	+	+
sNPF	XP_022672325.1 XP_022670836.1	ND*	XP_022700525.1 XP_022700540.1	ND*	+	+	+	+
SIFamide	XP_022668461.1 XP_022649195.1	LOC111250327	XP_022708258.1 XP_022702757.1	LOC111264575	+	+	+	+
SK	XP_022671216.1	LOC111244598	XP_022711355.1	SRR3635001.36448511.1	+	+	+	ND*
Tachykinin	XP_022673516.1	LOC111250114	XP_022703771.1	LOC111265999	+	+	+	+
Trissin	XP_022659797.1	SRR8864012.95461314.2	XP_022696460.1	LOC111258892	+	+	+	+

- **Accession numbers for the are provided for receptor sequence of the *V. destructor* and *V. jacobsoni* reported in this study.**
- **Gene ID's and SRA accession numbers are provided for ligands/neuropeptide gene sequence for *V. destructor* and *V. jacobsoni.***
- **Receptor sequences for *D. melanogaster* and *A. mellifera* were taken from Hauser et al., 2006.**
- **Receptor sequences for *T. urticae* were taken from Veenstra et al., 2012.**
- **ND** is **either Not Found (ND*) or Unknown (ND^).**

GPCR sequences:

ACP receptor:

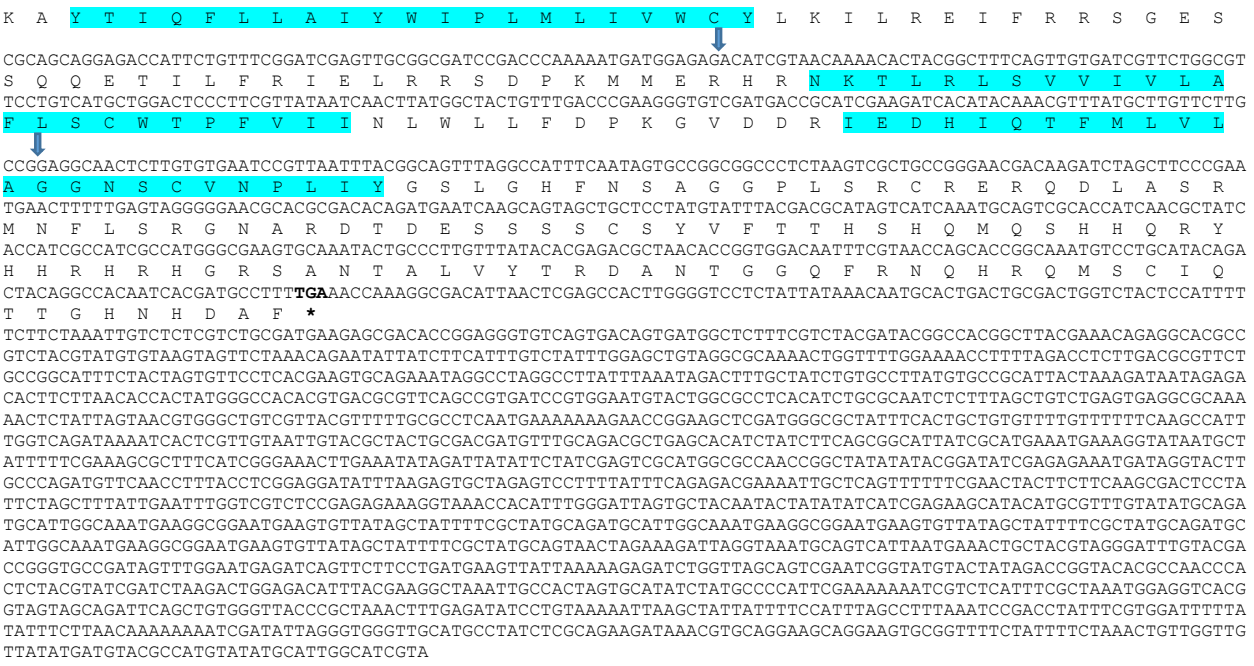
A) Gene structure:

V. destructor ACP receptor 1:

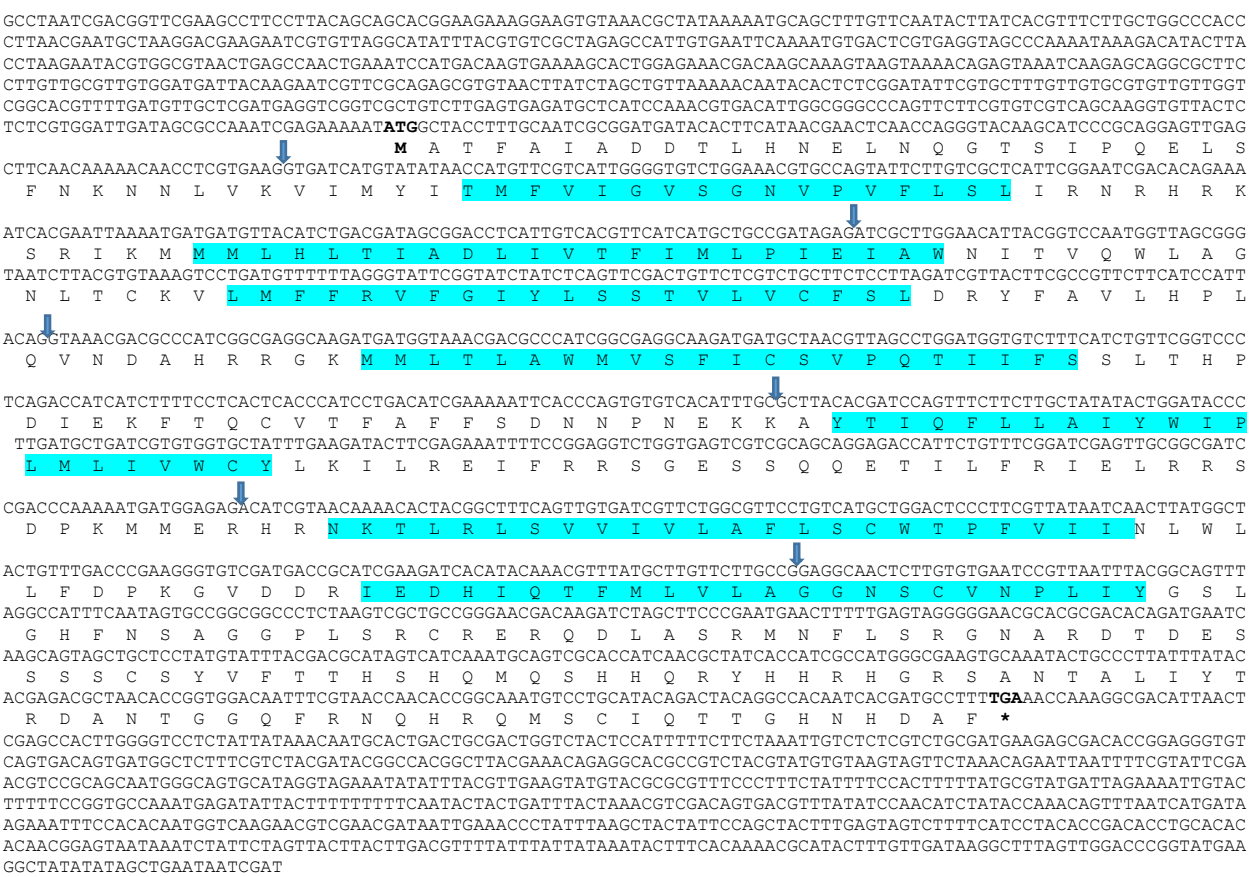
```
CTGGTTTCCAATGAAGATGTCTTTTTAATAACCAATTTTCCGGCAAATTCTGGCGTGACAACCATGCCTATATCGTTTGATGCTTCATATTTATTTGAGTTGGGCATTAGC
CAGATCTGTTAAATTAACTTGAAGATTTAGGTATATATTTTTTTGCAAGTGATTCTCAAAAAATTACCATTTAATGCTAGCAGCGATCCAAGTGTGTAACGATTAAATTGGA
AGTTGAAAACAAACTCTGCTGCTAAATTCGAATCTGGAATTCGTCGATAATCTCAAAAAAAGGACGTGCTTAGCCCGGGGCCGGATCACACATTTTTTAACGTCTTCACAT
AAAAGCGAAATGAAGTCAGATTATTCCTACTATGAGGCATAACAAGTTGAAGCTGAATCGTAGAGTTTGTGCCGTTATTGAGTACACCTAATGTCTGGATAGTAATATAGT
GCGGTAGATGTAACACTGAATTTAAGCGCACAGATCTGTGTACAGCATCTTCAAGACCAATGAGCGTCTTTGGATCTGTCAAAAGCGCCATTAAGCTACAATCATGAAACA
                                                                                                          M  K  Q
AATGGCGAATGAAGATACTGTTGGAAACCAAAGCATTTTAGTACAGCCATGGCCCTCAATCGACGATCTTTACCCGCTCGAAGAGGCCGAGGAATCGACGTCGCTGTGTCC
  M  A  N  E  D  T  V  G  N  Q  S  I  L  V  Q  P  W  P  S  I  D  D  L  Y  P  L  E  E  A  E  E  S  T  S  L  C  P
GGAGTGTCAATTTAATTCAGGGTCTCTTGTGAGCATTCTGGTCTACTCCGCGCTTTTTGTTTGCGCCACTGTGGGGAACGTACCTGTCTTTCTATCCCTGATCAGGAACAG
  E  C  Q  F  N  S  G  S  L  V  S  I  L  V  Y  S  A  L  F  V  C  A  T  V  G  N  V  P  V  F  L  S  L  I  R  N  R

GCACCGAAAGTCACGCATTAAACTTATGATGCTTAATCTCGCGATCGCCGATCTCATCGTAACCTTCATTTTCATTCCCGTCGAGATTTTTTGGCGCCTAATGGTTCAGTG
  H  R  K  S  R  I  K  L  M  M  L  N  L  A  I  A  D  L  I  V  T  F  I  F  I  P  V  E  I  F  W  R  L  M  V  Q  W

GAAAGCAGGCGTGTTAATGTGCAAAGTGTTTCAAGTGGTCCGGGCTTTTGGGCCATACCTCAGCTCAACAATCATCATATGCATTTCACTCGACCGATATTTTGCTGTTCT
  K  A  G  V  L  M  C  K  V  F  Q  V  V  R  A  F  G  P  Y  L  S  S  T  I  I  I  C  I  S  L  D  R  Y  F  A  V  L

TCATCCTCTAAAGGTTCACGATGCCCAGCGTCGTGGCAAAATTATGATCTTTTTAGCCTGGTTCGTGTCATTGACGTGGAGCTTACCACAGGCCGTGGTCTATGAGGTGGA
  H  P  L  K  V  H  D  A  Q  R  R  G  K  I  M  I  F  L  A  W  F  V  S  L  T  W  S  L  P  Q  A  V  V  Y  E  V  E

AGGTCATCCTTACATTAAGGACTTCTACCAGTGTGTGACGTTCAACTTCTTTACGAATGCGCCACAGCAAAGGGTATACGTTTTGTTCGGCCTGCTTCTTCAATATGGAAT
  G  H  P  Y  I  K  D  F  Y  Q  C  V  T  F  N  F  F  T  N  A  P  Q  Q  R  V  Y  V  L  F  G  L  L  L  Q  Y  G  I

CCCGTTGGTGATTATTATCTGGTGTTACTGCAAGATTCTTCTCGAGATTTGTCATCGATCAACAGACGGTGAGCTTACAACGAGGACTGGTCGACCCACAAATCTTCGAGT
  P  L  V  I  I  I  W  C  Y  C  K  I  L  L  E  I  C  H  R  S  T  D  G  E  L  T  T  R  T  G  R  P  T  N  L  R  V

GCAGTATAGGCAAGTCGTGCGACTGCGGCGTTCAGTTCCGAGCAACGTCGAGTGTATTACGCGAACACGCAACCGCACTCTACGGCTGACCGTTATCATCGTACTTACGTT
  Q  Y  R  Q  V  V  R  L  R  R  S  V  P  S  N  V  E  C  I  T  R  T  R  N  R  T  L  R  L  T  V  I  I  V  L  T  F

CTTCTGGTGCTGGACACCGTATGTCACAATGGTAGTCTGGTACCAAATCGATACCGAAGGGGCCACCAAGCTCAACGAGTATCTGCAAAACGCCCTGTTCATGTTTGCTGT
  F  W  C  W  T  P  Y  V  T  M  V  V  W  Y  Q  I  D  T  E  G  A  T  K  L  N  E  Y  L  Q  N  A  L  F  M  F  A  V

AAGCAATTCCTGCGTTAATCCGTTACTGTATGGCAGCTACGCTAAATCAAACTGGACGCGATTCTTTGCTAGATGTCTGGGTATCGGTGCCCACAAAGATGGGGCTGACGG
  S  N  S  C  V  N  P  L  L  Y  G  S  Y  A  K  S  N  W  T  R  F  F  A  R  C  L  G  I  G  A  H  K  D  G  A  D  G

CGGTTCTGTAGCCGGGGAACCCGGTCAACTTGAAAGAAACATTTCTTCACATTTCCGTCAAACGATCATCGTTACAAGTCGTGGCAGCTCTCATCGGAGTAAAATACTTTA
  G  S  V  A  G  E  P  G  Q  L  E  R  N  I  S  S  H  F  R  Q  T  I  I  V  T  S  R  G  S  S  H  R  S  K  I  L  Y
CTCGACCGTCGATCCGCCTTCTCCACTACAGCCAACATCAAGTATCCATTGTCACCAATCAGCAAACAGTCATCTTGATTAGAGCAAAAACAAGTTTTCTAGAAGTTTCGA
  S  T  V  D  P  P  S  P  L  Q  P  T  S  S  I  H  C  H  Q  S  A  N  S  H  L  D  *
AGCAAAAAACACCAGCAGCTTCCTTCGGGGCACCCAAAAAGGTGTAACAACGATACTAATTTTCAGTCGGTCTCAACAATACTCCATGTACATAGACATATAATATTATGG
AAGTTATAAAATGTCTCCGAACGTTGCGCCAACTGAAAACGATTACTGAAAAATATGTATCGCAATAAATACCAAAAACAGTTCCACCTTTATATATGTAAACTAGAACGA
TCGAGGTAAAATAATAATAATAAAAATCTTTATATATATATATATATATATATATGGCTTATGTTTGAACAACTTTCCTATTCTAGATTGAAAAACATTTCGGTAGTGGAA
TAATGTTAATTTTTGATTTTCGTTTAACCTAAGAATAGCACCATGGTGCCACGGAGTCGCCTTAACTGCAGGTTCATTATATTTGAGCTTTTAATCGTATAGAAGCATAGT
GCAGGAAAAACTAGAAATCAATGTGTTGAGTGCCATCTGAATTATTTTACTACTTCTTGTGTATAATTTACGACTGCATGCAACATTCTATTTCATATATGAGGCAGTTCT
TAACATATAGAGAAGCTATGAGCCGTTAAGTATATA
```

V. destructor ACP receptor 2:

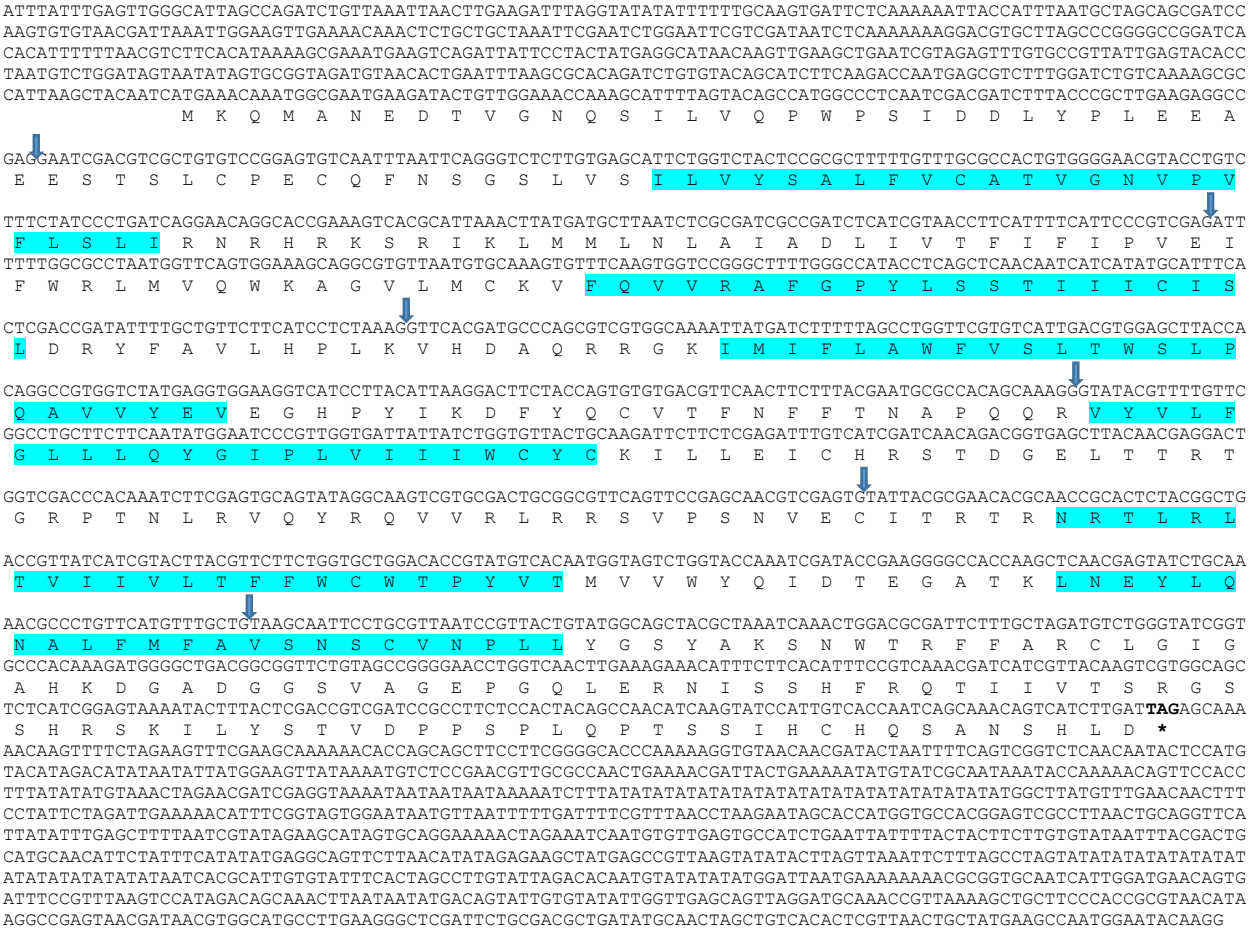
```
GTTAGTCACCCTTTGTCACCTTTACAGCCTGGTCTCCGCCCAGCGTACCGATCGCATCCGTATGATATTTTCCGCGTGACTCAGGATCGCTTGGACTTATTTGGATAACCC
AATTATCAGGCTTCAGAGAATCAGGCACTCGGCCAAATTGGAATCAAGCTGCTACGACTTGAAAATGAGTGATCATGTATATAACCATGTTCGTCATTGGGGTGTCTGGAA
                                                                                  M  Y  I  T  M  F  V  I  G  V  S  G

ACGTGCCAGTATTCTTGTCGCTCATTCGGAATCGACACAGAAAATCACGAATTAAAATGATGATGTTACATCTGACGATAGCGGACCTCATTGTCACGTTCATCATGCTGC
N  V  P  V  F  L  S  L  I  R  N  R  H  R  K  S  R  I  K  M  M  M  L  H  L  T  I  A  D  L  I  V  T  F  I  M  L

CGATAGAGATCGCTTGGAACATTACGGTCCAATGGTTAGCGGGTAATCTTACGTGTAAAGTCCTGATGTTTTTTAGGGTATTCGGTATCTATCTCAGTTCGACTGTTCTCG
P  I  E  I  A  W  N  I  T  V  Q  W  L  A  G  N  L  T  C  K  V  L  M  F  F  R  V  F  G  I  Y  L  S  S  T  V  L

TCTGCTTCTCCTTAGATCGTTACTTCGCCGTTCTTCATCCATTACAGGTAAACGACGCCCATCGGCGAGGCAAGATGATGCTAACGTTAGCCTGGATGGTGTCTTTCATCT
V  C  F  S  L  D  R  Y  F  A  V  L  H  P  L  Q  V  N  D  A  H  R  R  G  K  M  M  L  T  L  A  W  M  V  S  F  I
GTTCGGTCCCTCAGACCATCATCTTTTCCTCACTCACCCATCCTGACATCGAAAAATTCACCCAGTGTGTCACATTTGCGTTCTTCTCTGACAACAACCCTAATGAAAAAA
C  S  V  P  Q  T  I  I  F  S  S  L  T  H  P  D  I  E  K  F  T  Q  C  V  T  F  A  F  F  S  D  N  N  P  N  E  K

AGGCTTACACGATCCAGTTTCTTCTTGCTATATACTGGATACCCTTGATGCTGATCGTGTGGTGCTATTTGAAGATACTTCGAGAAATCTTCCGGAGGTCTGGTGAGTCGT
```



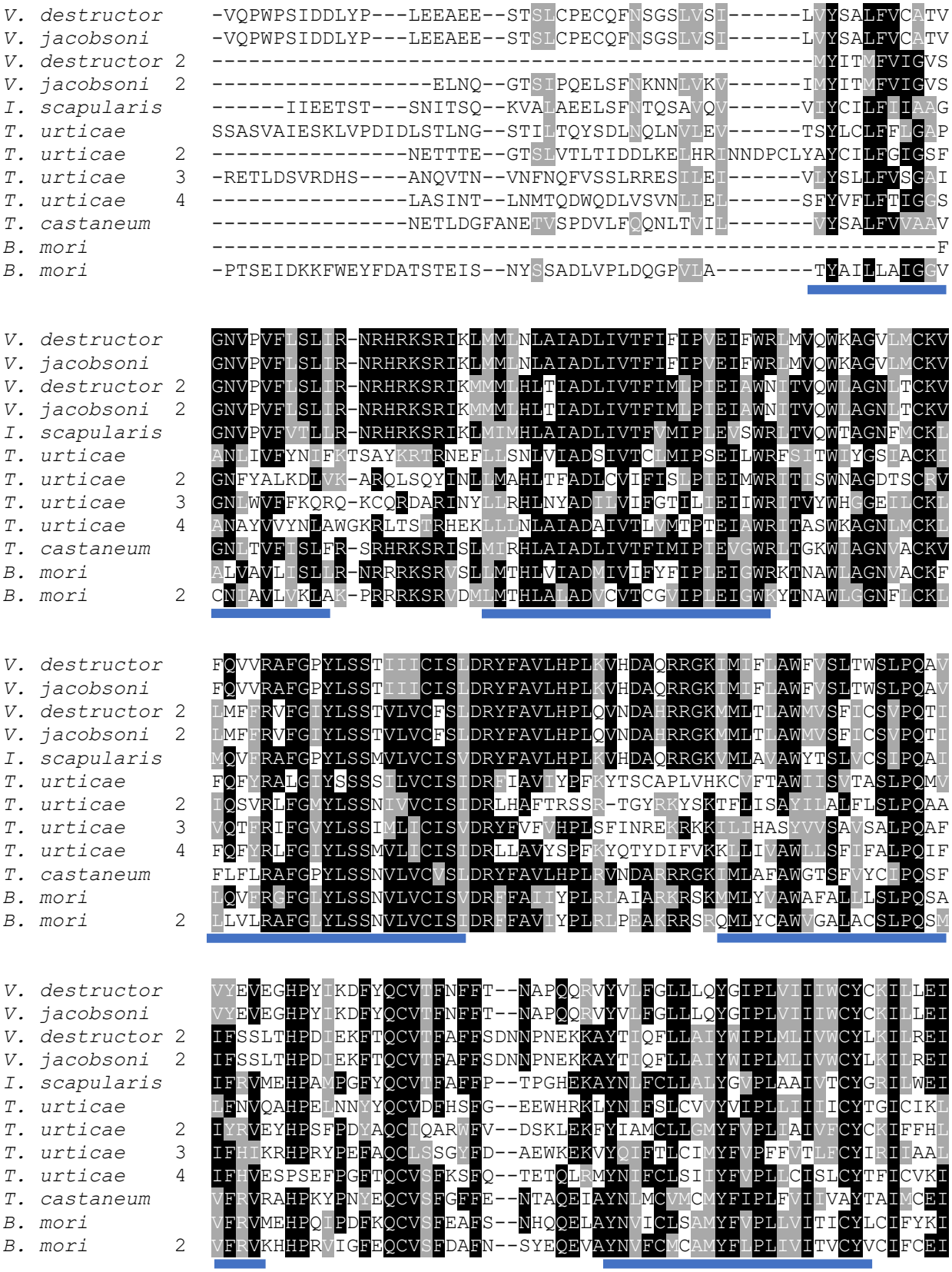
V. jacobsoni ACP receptor1:



V. jacobsoni ACP receptor2:


ATTTATTTGAGTTGGGCATTAGCCAGATCTGTTAAATTAACTTGAAGATTTAGGTATATATTTTTTGCAAGTGATTCTCAAAAAATTACCATTTAATGCTAGCAGCGATCC
AAGTGTGTAACGATTAAATTGGAAGTTGAAAACAAACTCTGCTGCTAAATTCGAATCTGGAATTCGTCGATAATCTCAAAAAAAGGACGTGCTTAGCCCGGGGCCGGATCA
CACATTTTTTAACGTCTTCACATAAAAGCGAAATGAAGTCAGATTATTCCTACTATGAGGCATAACAAGTTGAAGCTGAATCGTAGAGTTTGTGCCGTTATTGAGTACACC
TAATGTCTGGATAGTAATATAGTGCGGTAGATGTAACACTGAATTTAAGCGCACAGATCTGTGTACAGCATCTTCAAGACCAATGAGCGTCTTTGGATCTGTCAAAAGCGC
CATTAAGCTACAATCATGAAACAAATGGCGAATGAAGATACTGTTGGAAACCAAAGCATTTTAGTACAGCCATGGCCCTCAATCGACGATCTTTACCCGCTTGAAGAGGCC
M K Q M A N E D T V G N Q S I L V Q P W P S I D D L Y P L E E A
GAGGAATCGACGTCGCTGTGTCCGGAGTGTCAATTTAATTCAGGGTCTCTTGTGAGCATTCTGGTCTACTCCGCGCTTTTTGTTTGCGCCACTGTGGGGAACGTACCTGTC
E E S T S L C P E C Q F N S G S L V S I L V Y S A L F V C A T V G N V P V
TTTCTATCCCTGATCAGGAACAGGCACCGAAAGTCACGCATTAAACTTATGATGCTTAATCTCGCGATCGCCGATCTCATCGTAACCTTCATTTTCATTCCCGTCGAGATT
F L S L I R N R H R K S R I K L M M L N L A I A D L I V T F I F I P V E I
TTTTGGCGCCTAATGGTTCAGTGGAAAGCAGGCGTGTTAATGTGCAAAGTGTTTCAAGTGGTCCGGGCTTTTGGGCCATACCTCAGCTCAACAATCATCATATGCATTTCA
F W R L M V Q W K A G V L M C K V F Q V V R A F G P Y L S S T I I I C I S
CTCGACCGATATTTTGCTGTTCTTCATCCTCTAAAGGTTCACGATGCCCAGCGTCGTGGCAAAATTATGATCTTTTTAGCCTGGTTCGTGTCATTGACGTGGAGCTTACCA
L D R Y F A V L H P L K V H D A Q R R G K I M I F L A W F V S L T W S L P
CAGGCCGTGGTCTATGAGGTGGAAGGTCATCCTTACATTAAGGACTTCTACCAGTGTGTGACGTTCAACTTCTTTACGAATGCGCCACAGCAAAGGGTATACGTTTTGTTC
Q A V V Y E V E G H P Y I K D F Y Q C V T F N F F T N A P Q Q R V Y V L F
GGCCTGCTTCTTCAATATGGAATCCCGTTGGTGATTATTATCTGGTGTTACTGCAAGATTCTTCTCGAGATTTGTCATCGATCAACAGACGGTGAGCTTACAACGAGGACT
G L L L Q Y G I P L V I I I W C Y C K I L L E I C H R S T D G E L T T R T
GGTCGACCCACAAATCTTCGAGTGCAGTATAGGCAAGTCGTGCGACTGCGGCGTTCAGTTCCGAGCAACGTCGAGTGTATTACGCGAACACGCAACCGCACTCTACGGCTG
G R P T N L R V Q Y R Q V V R L R R S V P S N V E C I T R T R N R T L R L
ACCGTTATCATCGTACTTACGTTCTTCTGGTGCTGGACACCGTATGTCACAATGGTAGTCTGGTACCAAATCGATACCGAAGGGGCCACCAAGCTCAACGAGTATCTGCAA
T V I I V L T F F W C W T P Y V T M V V W Y Q I D T E G A T K L N E Y L Q
AACGCCCTGTTCATGTTTGCTGTAAGCAATTCCTGCGTTAATCCGTTACTGTATGGCAGCTACGCTAAATCAAACTGGACGCGATTCTTTGCTAGATGTCTGGGTATCGGT
N A L F M F A V S N S C V N P L L Y G S Y A K S N W T R F F A R C L G I G
GCCCACAAAGATGGGGCTGACGGCGGTTCTGTAGCCGGGGAACCTGGTCAACTTGAAAGAAACATTTCTTCACATTTCCGTCAAACGATCATCGTTACAAGTCGTGGCAGC
A H K D G A D G G S V A G E P G Q L E R N I S S H F R Q T I I V T S R G S
TCTCATCGGAGTAAAATACTTTACTCGACCGTCGATCCGCCTTCTCCACTACAGCCAACATCAAGTATCCATTGTCACCAATCAGCAAACAGTCATCTTGATTAGAGCAAA
S H R S K I L Y S T V D P P S P L Q P T S S I H C H Q S A N S H L D *
AACAAGTTTTCTAGAAGTTTCGAAGCAAAAAACACCAGCAGCTTCCTTCGGGGCACCCAAAAAGGTGTAACAACGATACTAATTTTCAGTCGGTCTCAACAATACTCCATG
TACATAGACATATAATATTATGGAAGTTATAAAATGTCTCCGAACGTTGCGCCAACTGAAAACGATTACTGAAAAATATGTATCGCAATAAATACCAAAAACAGTTCCACC
TTTATATATGTAAACTAGAACGATCGAGGTAAAATAATAATAATAAAAATCTTTATATAT

B) Multiple sequence alignment:



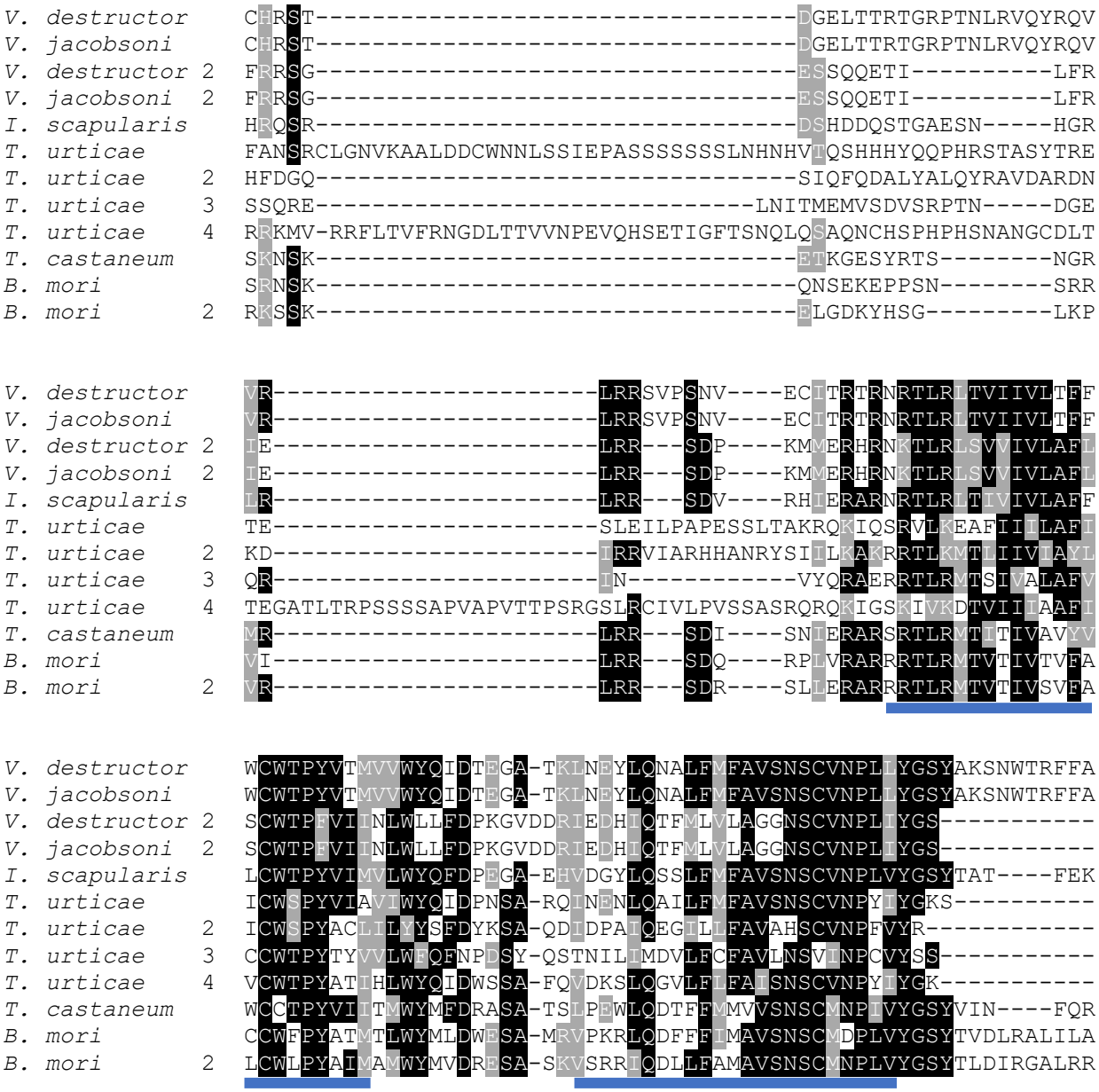
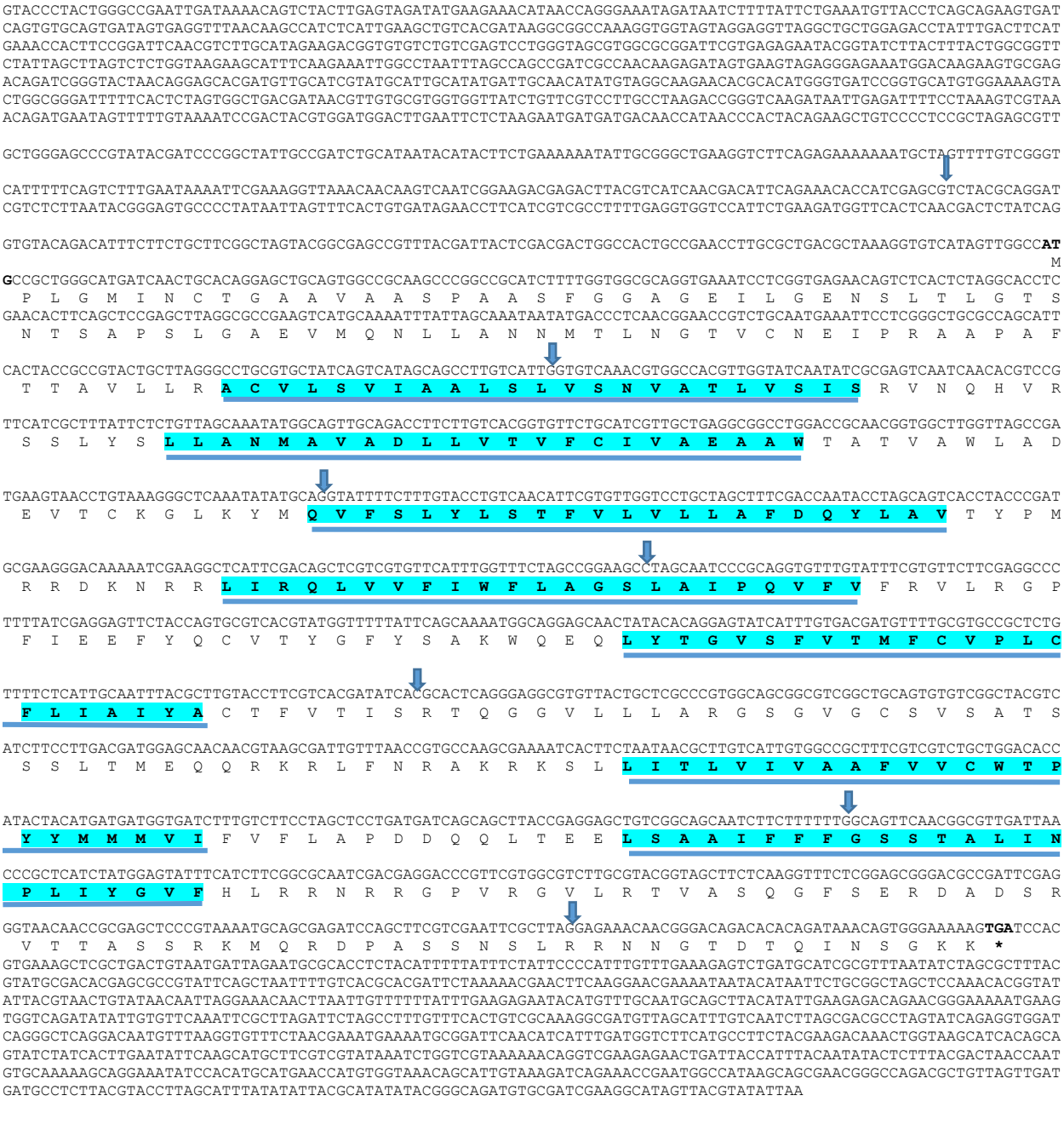


Figure 3.8 A) Gene structure for ACP receptor sequence in *Varroa* mite is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of ACP receptor from various orthologous species across Phylum Arthropoda

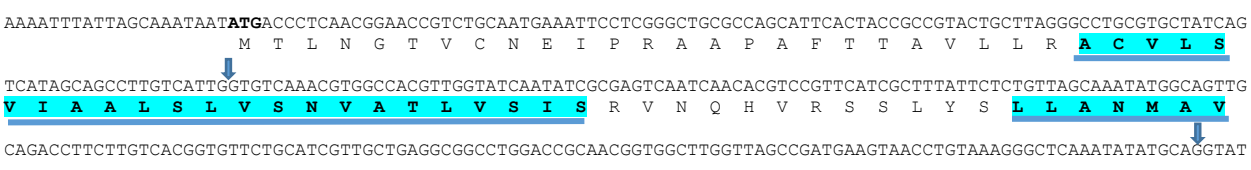
Corazonin receptor

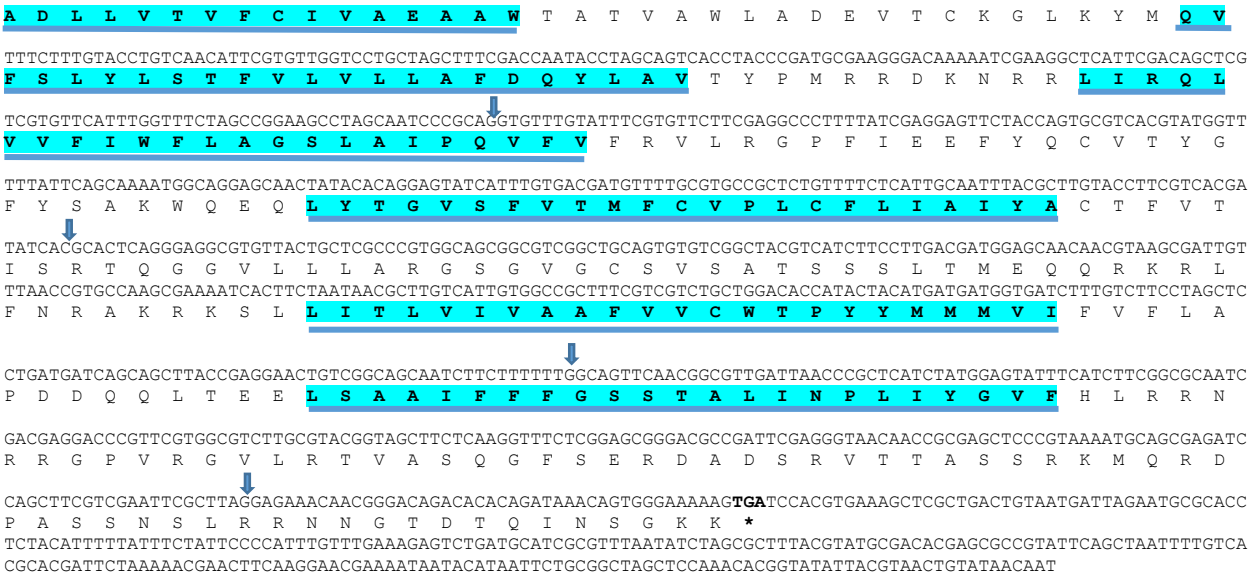
A) Gene structure:

V. destructor

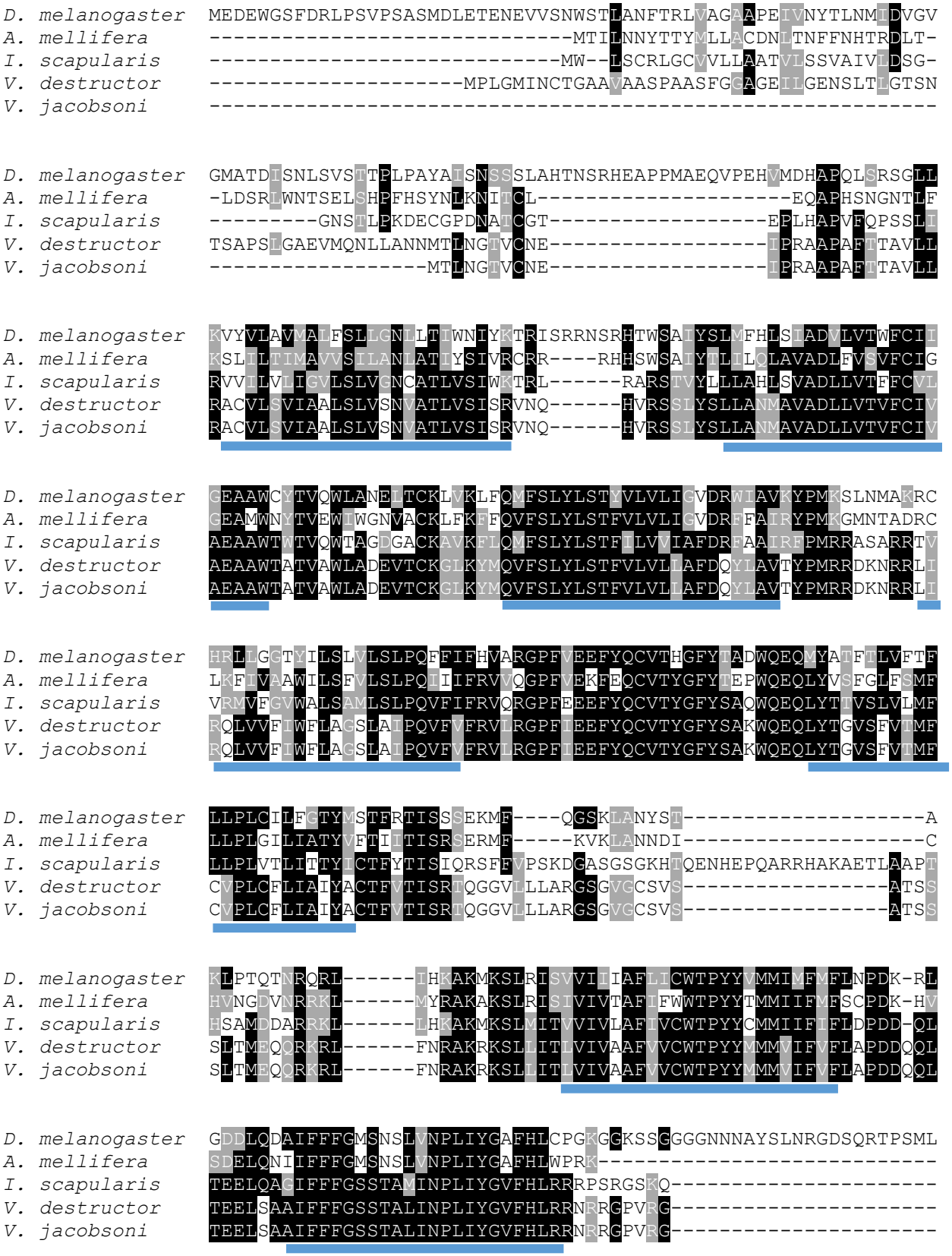


V. jacobsoni




A D L L V T V F C I V A E A A W T A T V A W L A D E V T C K G L K Y M Q V
TTTCTTTGTACCTGTCAACATTCGTGTTGGTCCTGCTAGCTTTCGACCAATACCTAGCAGTCACCTACCCGATGCGAAGGGACAAAAATCGAAGGCTCATTCGACAGCTCG
F S L Y L S T F V L V L L A F D Q Y L A V T Y P M R R D K N R R L I R Q L
TCGTGTTCATTTGGTTTCTAGCCGGAAGCCTAGCAATCCCGCAGGTGTTTGTATTTCGTGTTCTTCGAGGCCCTTTTATCGAGGAGTTCTACCAGTGCGTCACGTATGGTT
V V F I W F L A G S L A I P Q V F V F R V L R G P F I E E F Y Q C V T Y G
TTTATTCAGCAAAATGGCAGGAGCAACTATACACAGGAGTATCATTTGTGACGATGTTTTGCGTGCCGCTCTGTTTTCTCATTGCAATTTACGCTTGTACCTTCGTCACGA
F Y S A K W Q E Q L Y T G V S F V T M F C V P L C F L I A I Y A C T F V T
TATCACGCACTCAGGGAGGCGTGTTACTGCTCGCCCGTGGCAGCGGCGTCGGCTGCAGTGTGTCGGCTACGTCATCTTCCTTGACGATGGAGCAACAACGTAAGCGATTGT
I S R T Q G G V L L L A R G S G V G C S V S A T S S S L T M E Q Q R K R L
TTAACCGTGCCAAGCGAAAATCACTTCTAATAACGCTTGTCATTGTGGCCGCTTTCGTCGTCTGCTGGACACCATACTACATGATGATGGTGATCTTTGTCTTCCTAGCTC
F N R A K R K S L L I T L V I V A A F V V C W T P Y Y M M M V I F V F L A
CTGATGATCAGCAGCTTACCGAGGAACTGTCGGCAGCAATCTTCTTTTTTGGCAGTTCAACGGCGTTGATTAACCCGCTCATCTATGGAGTATTTCATCTTCGGCGCAATC
P D D Q Q L T E E L S A A I F F F G S S T A L I N P L I Y G V F H L R R N
GACGAGGACCCGTTCGTGGCGTCTTGCGTACGGTAGCTTCTCAAGGTTTCTCGGAGCGGGACGCCGATTCGAGGGTAACAACCGCGAGCTCCCGTAAAATGCAGCGAGATC
R R G P V R G V L R T V A S Q G F S E R D A D S R V T T A S S R K M Q R D
CAGCTTCGTCGAATTCGCTTAGGAGAAACAACGGGACAGACACACAGATAAACAGTGGGAAAAAGTGATCCACGTGAAAGCTCGCTGACTGTAATGATTAGAATGCGCACC
P A S S N S L R R N N N G T D T Q I N S G K K *
TCTACATTTTTATTTCTATTCCCCATTTGTTTGAAAGAGTCTGATGCATCGCGTTTAATATCTAGCGCTTTACGTATGCGACACGAGCGCCGTATTCAGCTAATTTTGTCA
CGCACGATTCTAAAAACGAACTTCAAGGAACGAAAATAATACATAATTCTGCGGCTAGCTCCAAACACGGTATATTACGTAACTGTATAACAAT

B) Multiple sequence alignment:



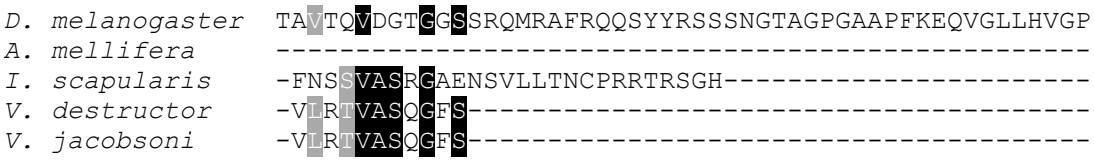
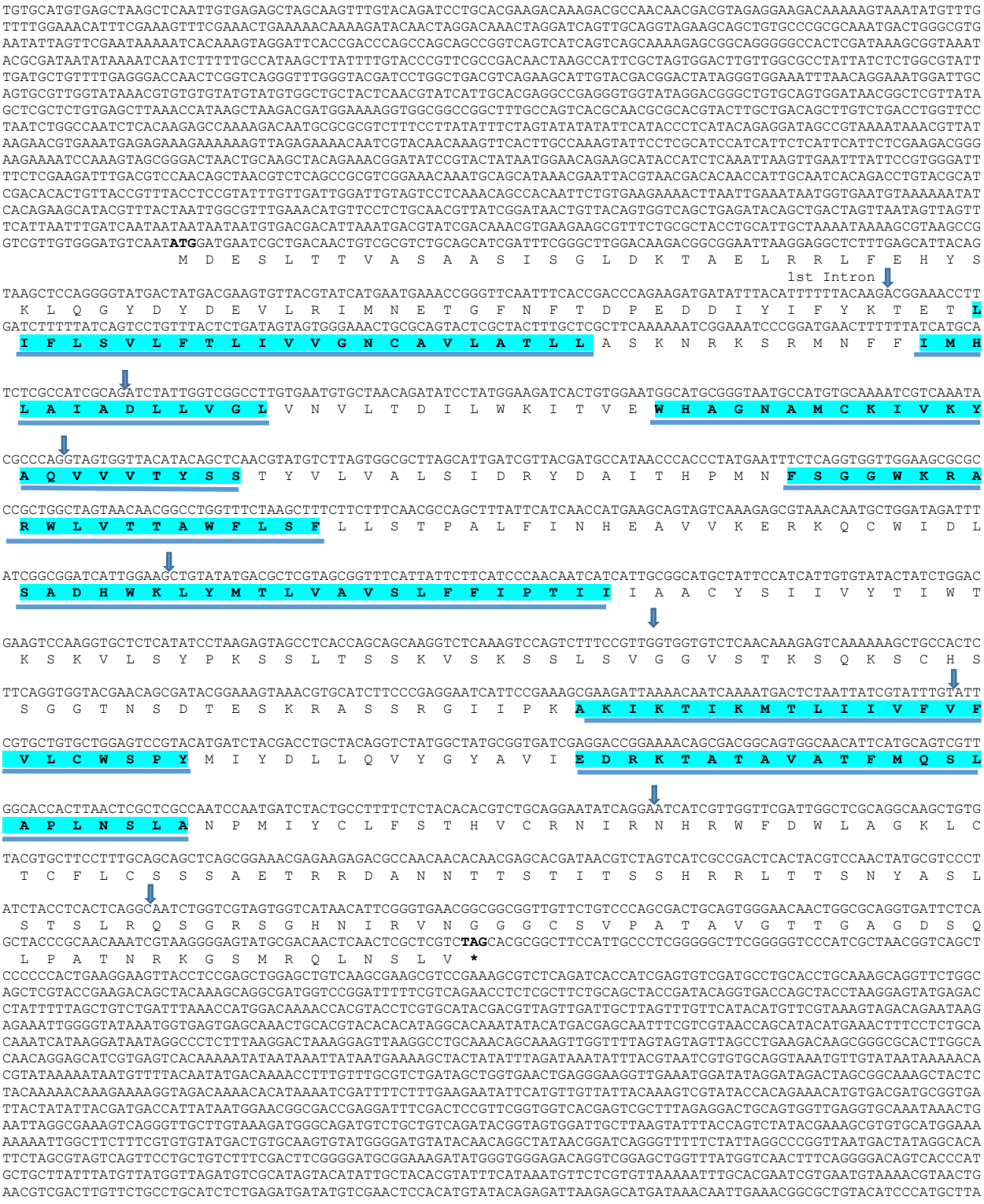


Figure 3.9 A) Gene structure for Corazonin receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of Corazonin receptor from various orthologous species across Phylum Arthropoda.

CCAP receptor:

A) Gene structure:

V. destructor



```
TTCTATTATGTAAACATGAAAAAATACAGGTTCACGTACAAATATTGTGTTATTTATTATAGTAATATAATAAGACCTTCCTGAAATGACTTGTTCAGGATTTGTGTTTTC
CACTATGAAAAAGCGTAAAGTTTTATTTAAAAATGTAAACCAAATTCACACAGCAAAATGTAAATAAATAGCGTATTTATTTTTTATAAAACTGAAGCAAGCTTTCCGAGA
ATGTTTATTTAACCAGTGAGCGCCCATCGAAGCTATTCAAAATTGCATTCAACTTTCTGTTCGGTTTCCATCGACTGGAACTCGACGACCCGATGATCTTCGGATCTAGAA
GCAAATTAAATAACTAACAATTGACGAAATTAAAGTATCGATCATATTTGGTTTAGTTTGAATGAGGTAGTGGTGTGGGCGCAAAGAATCGTGCGCTAATCCTTGATGGCC
TAGAATGTTTTTAAAAATAAGCTGAATTTTCTTTTTTCCGACCAGCTGAGGGATTAATTTTAGAGAAAAATTACGTAGAACATAATGTCGAACAGAACTTTTACTCGTTAT
ATATGCGTTTTATGTTATAAACA
```

B) Multiple sequence alignment:

```
D. melanogaster  MLHLRLFDSSLYYTLASASESSGLASSTSTERSFNGT----------------------
A. mellifera     MS------VEE---------------------IVNRT----------------------
I. scapularis    MAT------EV-TAVDGT-------------------------------------------
T. urticae       MTM------FN-TTFPSFPS-TDLQSTSPVDIVYRFNDQIVDFNSLHHQRSSPSPHSFQP
V. destructor    MDE------SL-TTVASAASISGLDKTAEIRRLFEHY----------------------
V. jacobsoni     MDE------SL-TTVASAASISGLDKTAEIRRLFEHY----------------------

D. melanogaster  ----QGAG--GVAAGGESLTPTDVAAVNLTYFTPAISHVMLAPTTIATTTASATMVQIQT
A. mellifera     ----DDNH--TFQKEGSTWNFTDYVDT-------------------------------
I. scapularis    ------------------------------------M---------------------
T. urticae       SQHSSKLLLNGT-----------A---------DWSTNGS--------------------
V. destructor    ----SKLQ--GYD----------Y---------DEVLRIM--------------------
V. jacobsoni     ----SKLQ--GYD----------Y---------DEVLRIM--------------------

D. melanogaster  TAAPSHDLETGGNST-SSDPGEFDNLNSFYFYETEQFAVLWILFTVIVLGNSAVLFVMFI
A. mellifera     -------NVTQLDAN-A------TKIDSFYFYKTEQFTVLWLLFAMIVVGNIAVLIGLQW
I. scapularis    -------NTTSGNATSE------PELELFYFYQAETLTFLWILFAMIVLGNSAVLVALTL
T. urticae       -------IEFTFNETTE------KDFQVYYFYQTEQLTFLWIIFVMIVFGNCSVLGTLLL
V. destructor    -------NETGFNFTD------PEDDIYIFYKTETLIFLSVLFTLIVVGNCAVLATLLA
V. jacobsoni     -------NETGFNFTD------PEDDIYIFYKTETLIFLSVLFTLIVVGNCAVLATLLA

D. melanogaster  NKNRKSRMNYFIKQLALADLCVGLLNVLTDIIWRITTSWRAGNLACKAIRFSQVCVTYSS
A. mellifera     GKRRKTRMDFFIKQLAFADLLVGLISVLTDIIWKTTVSWHAGNVACKLIRFMQAVVTYSS
I. scapularis    SKSRKSRMNFFIMHLAIADLAVGLINVLTDIVWKTTVDWYGGNVGCKLVKFAQILVTYSS
T. urticae       SKGRKSRMNFFIMHLAIADLLVGLINVLTDIVWRFTVGFYTGDTACKAIKFAQVVVTYGS
V. destructor    SKNRKSRMNFFIMHLAIADLLVGLVNVLTDILWKITVEWHAGNAMCKIVKYAQVVVTYSS
V. jacobsoni     SKNRKSRMNFFIMHLAIADLLVGLVNVLTDILWKITVEWHAGNAMCKIVKYAQVVVTYSS

D. melanogaster  TYVLVAMSIDRYDAITHPMNFSKSWKRARHLVAGAWLISALFSIPILVLYEEKLIQGHPQ
A. mellifera     TYVLVALSIDRYDAITRPMNFTGRWWRARALVISAWGLSALFSAPIIFLYEEKRVEGKTQ
I. scapularis    TYVLVALSIDRYDAITNPMNFSGSWKRARCLVGLAWGASSAMSVPAVFLSREALVRGRLQ
T. urticae       TYVLVALSIDRYDAITHPLNFTIRARRAKFLIAFAWTLSGIFSIPTLFLYKIHNIEDKPQ
V. destructor    TYVLVALSIDRYDAITHPMNFSGGWKRARWLVTTAWFLSFLLSTPALFINHEAVVKERKQ
V. jacobsoni     TYVLVALSIDRYDAITHPMNFSGGWKRARWLVTTAWFLSFLLSTPALFINHEAVVKERKQ

D. melanogaster  CWIELGSPIAWQVYMSLVSATLFAIPALIISACYAIIVKTIWAKGSIFVP-TER------
A. mellifera     CWIDLGSPTQWKIYMSLVSFTLFIAPTLIIGGCYAVIVATIWSQGGALRQ-G--------
I. scapularis    CWIELE-LWQWQLYMTLVACSLFFVPALVITACYSVIVYTIWTKSKILSY-PKL------
T. urticae       CWIHLE-PEEWQIYMTLVATSLFFIPTVIISACYSIIVHTIWTKSRLMGSSPRK------
V. destructor    CWIDLS-ADHWKLYMTLVAVSLFFIPTIIIAACYSIIVYTIWTKSKVLSY-PKSSLTSSK
V. jacobsoni     CWIDLS-ADHWKLYMTLVAVSLFFIPTIIIAACYSIIVYTIWTKSKVLSY-PKSSLTSSK

D. melanogaster  ------------------------AGFGAAPARRASSRGIIPRAKVKTVKMTLTIVFVFI
```

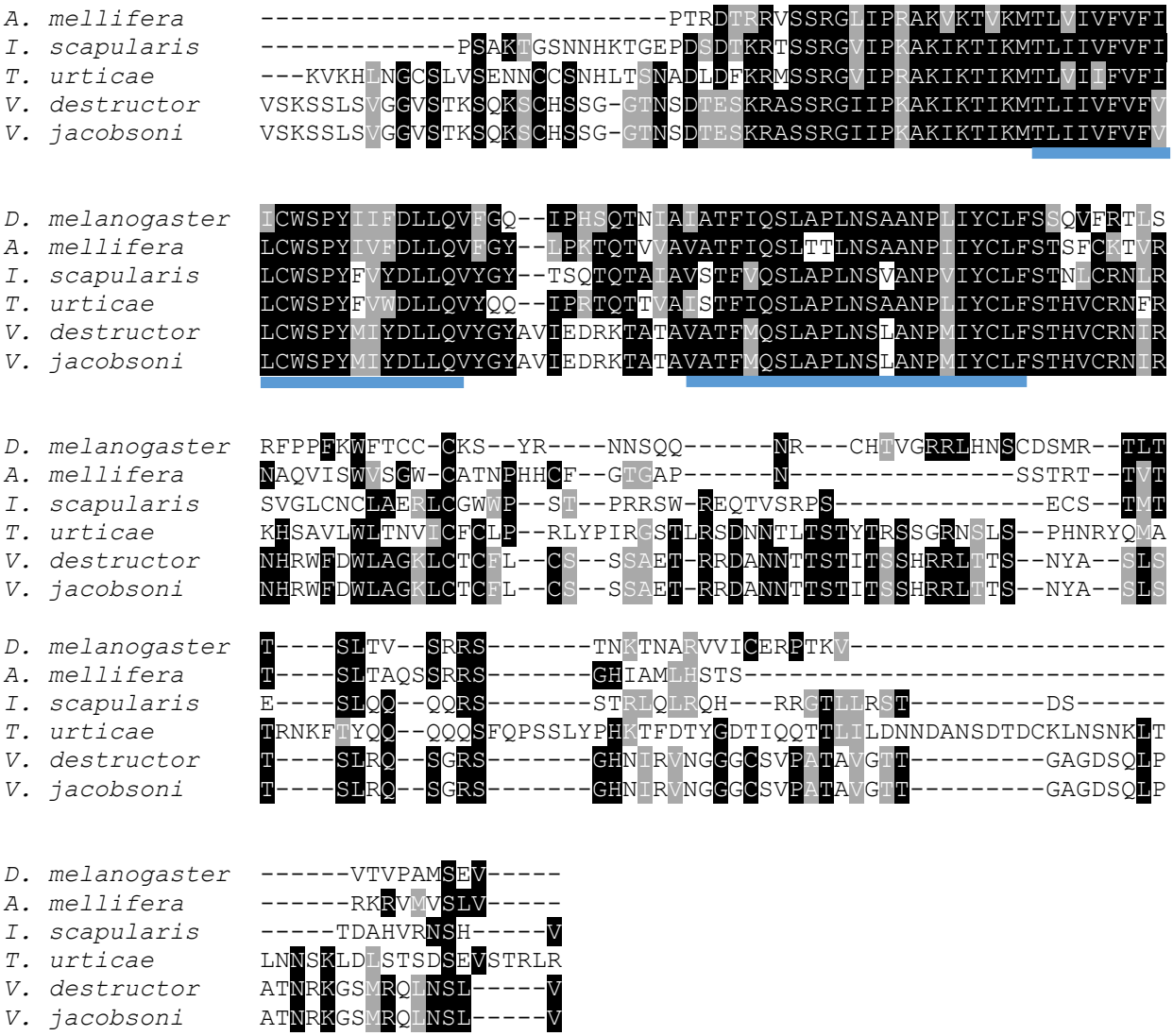
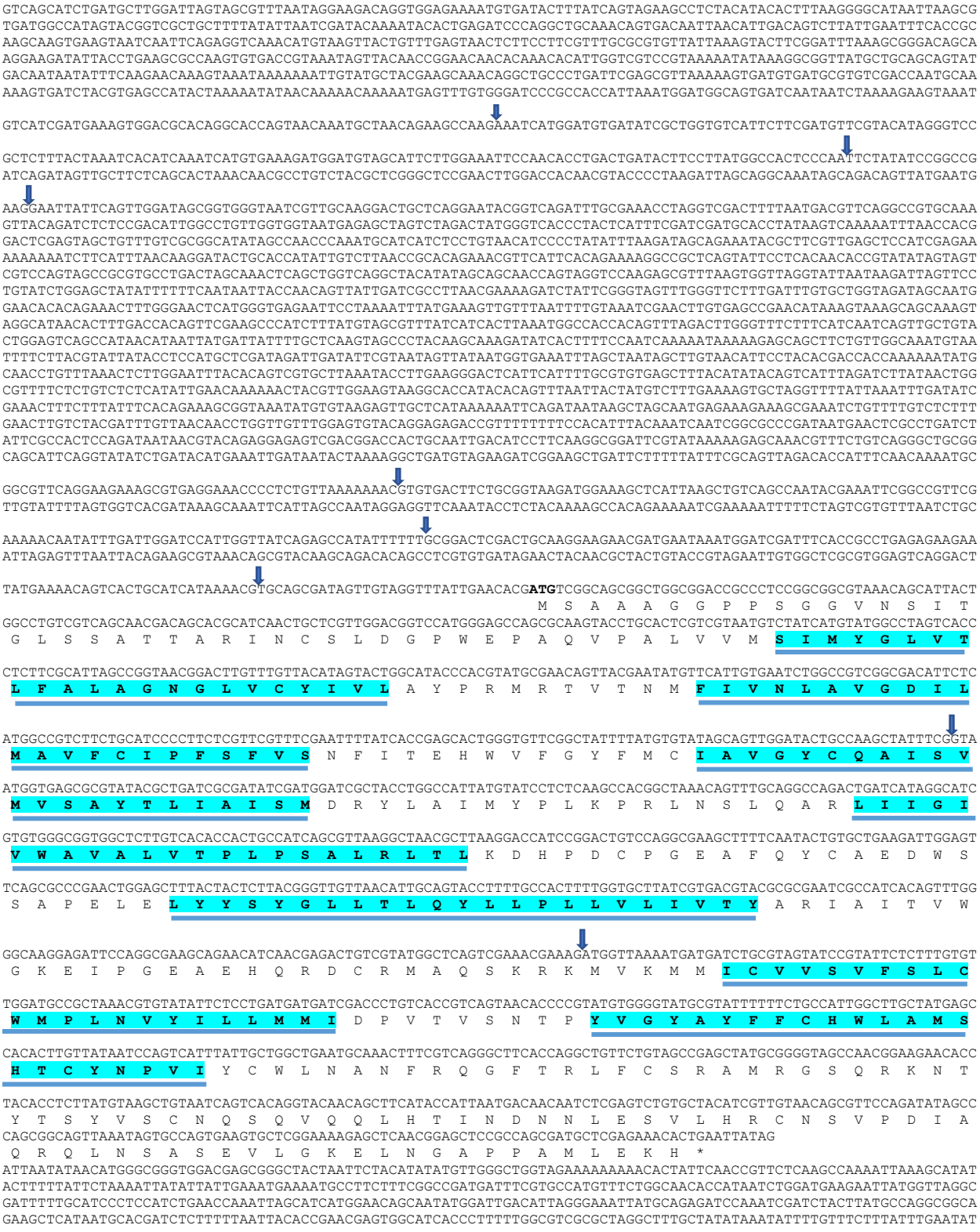


Figure 3.10 A) Gene structure for CCAP receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of CCAP receptor from various orthologous species across Phylum Arthropoda.

RYamide receptor

A) Gene structure:

V. destructor



```
TATTATTATTATTTTAAATGATTCAAATTTAAATGATCGACTGCTTTAGAGATGAGTCGATTTTGTCGCATTTCTATAGAGCTGAGTTGAAATGCGATCGCGTAAATTCAT
GTAGCATTCAAACATCGAGCTCTCTTTATAGCTAGCTTTTTTAAAATGCTTTAAAATAGTCTGGTGGTAAATGTTTTGTTTTTCGGCTATCTCTCTCATCGCTGTTCACGA
TTTTTCGTTCTTCTTGCTCAACTTTCGCATTAATATCATCAACTTTTTCTGTAGTCTGTAAACCACCACGAAAACCATCTTTCAGATCGAAATTCCTTGAAGGAAAACGAG
AGAACCATCTTCTTCTAGCAGCTTCTTCGAATAAAGTATTTTCACTAAGAACAGCCGTAGCCAGAACTTGTACCTTTATCGTAATAAACCTGTGAAATACAGCGAATTTTT
GCAGGAATGTTTCTCATCTCCGAGAGGCTTTATCTCAAGTCTGAATGAGGGCTCTCAAATGACACCAACGCCAATATGACGTAACTCGTATTACCTGAAGTCATCTAGCAA
GAATAGTGATCAATACTTCAACTTTTTTGACAACCCAATTTATATACACATACTTAACCAGTATACGACGACTGGTATTCTCCCTCTTAGAATCACCCATAGTGAAAATAA
TTATGTGTAGCTTATATTCGTTGGCCCGTACGGACGGTATAACTAATCTTTTTCTATAAAAGTGTTGGCTAATGGATTGATGCACGACCGGAACAGCCGAATCTATTTCTC
TTTATAGTACAATTTGTGTGGGCTTCATTGACCTGTTCGCAGCCGAATGAGACCGTACTCTGGGTGTGTAGGGCACGATAATACCCATTAAACTGCTGTATCTTTGGTGCT
TGCTGTAAGAAGCTATGAAACATTATTTCAGATTTTCCATGCTCTTTTTCAAAATTCTTCCTTTTTGCGCGTTCAGGTTCTTCTTTCTTGCTCGTACTTCATTTGATGCGC
ATAATACGTCGCATTTCCCTATCTTTGCGACTTAGGTCTGCTACTGTCCGTATTGTGGTTGGTGAGATCGCCATTTGTATTATCATGCCAATCCAGAGTGAAGAACCTCGT
TTAGTTCTTATCGTTCAAGACTGTCACCATTTTCACATTATTGCTGTTCTGATTCTGCAAATTTTC
```

V. jacobsoni

```
TCCTTTATCGAGCAACAATGGTTTTCTTTCGTCTCTCGTTCATTGCTTCGATGAACCGTCGTGTAGGCTTTACCCATCAAGGAATCTACATGTCTGAACAAGTTTTCTATC
ACACATTTATCCACGCCTACTTAAAGCACATGTGACAAACGAATAATAGAGAGAAATATATAACATGCTGCGCATCGAAAGCAGATGATCGTGCTAGCCTCACTCGGTGGC

CGTCGTTTAGCACGTAACGTAATAGTTGTGAGCGATCTCAGCAAGAAAATGCATATCATCGATGCAAAGATAAACCACAGGAACGACTACTGATCTGTTCGCAATCCAGAG

GATATGTTTATGAAGGATGGTAAAACAGAGTTCACTCCTTAACAAAAATCGACTTGGATCTAAAGGCATATCTGAGAAAATCCGTTTTGATCAAGCTTGTGGGTCACAGTA

ACTCGAAATCATGGATGTGATATCGCTGGTGTCATTCTTGGATGTTCGTACATAGGGTCCGCTCTTTACTAAATCACATCAAATCATGTGAAAGATGGATGTAGCATTCTT

GGAAATTCCAACACCTGACTGATACTTCCTTATGGCCACTCCCAAGTAAGCGATAAGCGACGTCCTGTGCAACACTGTTTGATGCATTTGTCACATGGGTCACGAAAAGTA

AGAGGAATCATTAGCGGAGAGGCCAAACCGCTTTTACGTATCATACGCTGCCTAATATTATTGTAAAATTAATTATTCTATATCCGGCCGATCAGACAGTTGCTTCTCAGC

ACTAAACAACGCCTGTCTACGCTCGGGCTCCGAACTTGGACCACAACGTACCCCTAAGATTAGCAGGCAAATAGCAGACAGTTATGAATGAAGGATGGATACCCAGTTCAA

CTTCAACAAAATCCAGTTCAACGAAAGCATTCTGAAGTGTGGCAAATTGAAATGCGTGCACGGAGTGTTAGGCGATAAAGGTGAAGACCGGTTCAGGAATTATTCAGTTGG
TGGGTAATCGTTGCAAGGACTGCTCAGGAATACGGTCAGATTTGCGAAACCTAGGTCGACTTTTAATGACGTTCAGGCCGTGCAAAGTTACAGATCTCTCCGACATTGGCC

TGTTGGTGATAATAACGTACAGAGGAGAGTCGACGGACCACTGCAATTGACATCCTTCAAGGCGGATTCTTATAAAAAGAGCAAACGTTTCTGTCAGGGCTGCGGCAGCAT

TCAGGCGGACTCGACTGCGAGGAAGAACGATGAATAAATGGATCGATTTCACCGCCTGAGAGAAGAAATTAGAGTTTAATTACAGAAGCGTAAACAGCGTACAAGCAGACA

CAGCCTCGTGGGATAGAACTACAACGCTACTGTACCGTAGAATTGTGGCTCGCGTGGAGTCAGGACTTATGAAAACAGTCACTGCATCATAAAACGACACGCCGATCAGCC

GGACCGATTATTGCAACTCACAAAATCTTGGGCCAGCCGAAGGGAATCAGCTTCCACTTAGGAAAC

GTGCAGCGATAGTTGTAGGTTTATTGAACACG ATG TCGGCAGCGGCTGGCGGACCGCCCTCCGGCGGCGTAAACAGCATTACTGGCCTGTCGTCAGCAACGACAGCACGCA
                                M S A A A G G P P S G G G V N S I T G L S S A T T A R
TCAACTGCTCGTTGGACGGTCCATGGGAGCCAGCGCAAGTACCTGCACTCGTCGTAATGTCTATCATGTATGGCCTAGTCACCCTCTTCGCATTAGCCGGTAACGGACTTG
I N C S L D G P W E P A Q V P A L V V M S I M Y G L V T L F A L A G N G L

TTTGTTACATAGTACTGGCATACCCACGTATGCGAACAGTTACGAATATGTTCATTGTGAATCTGGCCGTCGGCGACATTCTCATGGCCGTCTTCTGCATCCCCTTCTCGT
V C Y I V L A Y P R M R T V T N M F I V N L A V G D I L M A V F C I P F S

TCGTTTCGAATTTTATCACCGAGCACTGGGTGTTCGGCTATTTTATGTGTATAGCAGTTGGATACTGCCAAGCTATTTCGGTAATGGTGAGCGCGTATACGCTGATCGCGA
F V S N F I T E H W V F G Y F M C I A V G Y C Q A I S V M V S A Y T L I A

TATCGATGGATCGCTACCTGGCCATTATGTATCCTCTCAAGCCACGGCTAAACAGTTTGCAGGCCAGACTGATCATAGGCATCGTGTGGGCGGTGGCTCTTGTCACACCAC
I S M D R Y L A I M Y P L K P R L N S L Q A R L I I G I V W A V A L V T P

TGCCATCAGCGTTAAGGCTAACGCTTAAGGACCATCCGGACTGTCCAGGCGAAGCTTTTCAATACTGTGCTGAAGATTGGAGTTCAGCGCCCGAACTGGAGCTTTACTACT
L P S A L R L T L K D H P D C P G E A F Q Y C A E D W S S A P E L E L Y Y

CTTACGGGTTGTTAACATTGCAGTACCTTTTGCCACTTTTGGTGCTTATCGTGACGTACGCGCGAATCGCCATCACAGTTTGGGGCAAGGAGATTCCAGGCGAAGCAGAAC
S Y G L L T L Q Y L L P L L V L I V T Y A R I A I T V W G K E I P G E A E

ATCAACGAGACTGTCGTATGGCTCAGTCGAAACGAAAGATGGTTAAAATGATGATCTGCGTAGTATCCGTATTCTCTTTGTGTTGGATGCCGCTAAACGTGTATATTCTCC
H Q R D C R M A Q S K R K M V K M M I C V V S V F S L C W M P L N V Y I L

TGATGATGATCGACCCTGTCACCGTCAGTAACACCCCGTATGTGGGGTATGCGTATTTTTTCTGCCATTGGCTTGCTATGAGCCACACTTGTTATAATCCAGTCATTTATT
L M M I D P V T V S N T P Y V G Y A Y F F C H W L A M S H T C Y N P V I Y

GCTGGCTGAATGCAAACTTTCGTCAGGGTTTCACCAGGCTGTTCTGTAGCCGAGCTATGCGGGGTAGCCAACGGAAGAACACCTACACCTCTTATGTAAGCTGTAATCAGT
C W L N A N F R Q G F T R L F C S R A M R G S Q R K N T Y T S Y V S C N Q
CACAGGTACAACAGCTTCATACCATTAATGACAACAATCTCGAGTCTGTGCTACATCGTTGTAACAGCGTTCCAGATATAGACCAGCAGCAGTTAAATAGTGCCAGTGAAG
S Q V Q Q L H T I N D N N L E S V L H R C N S V P D I D Q Q Q L N S A S E
TGCTCGGAAAAGAGCTCAACGGAGCTCCGCCAGCGATGCTCGAGAAACAC TGA ATTATAGATTAATATAACATGGGCGGGTGGACGAGCGGGCTACTAATTCTACATATAT
V L G K E L N G A P P A M L E K H *
GTTGGGCTGGTAGAAAAAAAAACACTATTCAACCGTTCTCAAGCCAAAATTAAAGCATATACTTTTTATTCTAAAATTATATTATTGAAATGAAAATGCCTTCTTTCGGCC
GATGATTTCGTGCCATGTTTCTGGCAACACCATAATCTGGATGAAGAATTATGGTTAGGCGATTTTTGCATCCCTCCATCTGAACCAAATTAGCATCATGGAACAGCAATA
TGGATTGACATTAGGGAAATTATGCAGAGATCCAAATCGATCTACTTATGCCAGGCGGCAGAAGCTCATAATGCACGATCTCTTTTTAATTACACCGAACGAGTGGCATCA
CCCTTTTTGGCGTCGCGCTAGGCTTTGCTATATAAATATTTTGTTTCTTTATTTGAATATTATTATTATTTTAAATGATTCAAATTTAAATGATCGACTGCTTTAGAGATG
AGTCGATTTTGTCGCATTTCTATAGAGCTGAGTTGAAATGCGATCGCGTAAATTCATGTAGCATTCAAACATCGAGCTCTCTTTATAGCTAGCTTTTTTAAAATGCTTTAA
AATAGTCTGGTGGTAAATGTTTTGTTTTTCGGCTATCTCTCTCATCGCTGTTCACGATTTTTCGTTTTTCTTGCTCAACTTTCGCATTAATATCATCAACTTTTTCTGTAG
TCTGTAAACCACCACGAAAACCATCTTTCAGATCGAAATTCCTTGAAGGAAAACGAGAGAACCATCTTCTTCTAGCAGCTTCTTCGAATAAAGTATTTTCACTAAGAACAG
```

```
CCGTAGCTAGAACTTGTACCTTTATCGTAATAAACCTGTGAAATACAGCGAATTTTTGCAGGAATGTTTCTCATCTCCGAGAGGCTTTATCTCAAGTCTGAATGAGGGCTC
TCAAATGACACCAACGCCAATATGACGTAACTCGTATTACCTGAAGTCATCTAGCAAGAATAGTGATCAATACTTCAACTTTTTTGACAACCCAATTTATATACACATACT
TAACCAGTATACGACGACTGGTATTCTCCCTCTTAGAATCACCCATAGTGAAAATAATTATGTGTAGCTTATATTGGCCCGTACGGACGGTATAACTAATCTTTTTCTATA
AAAGTGTTGGCTAATGGATTGATGCACGACCGGAACAGCCGAAGCTATTTCTCTTTATAGTACAGTTTGTGTGGGCTTCATTGACCTGTTCGCAGCCGAATGAGACCATAC
TCTGGGTGTGTAGGGCACGATAATACCCATTAAACTGCTGTATCTTTGGTGCTTGCTGTAAGAAGCTATGAAACATTATTTCAGATTTTCCATGCTCTTTTTCAAAATTCT
TCCTTTTTGCGCGTTCAGGTTCTTCTTTCTTGCTCATACTTCATTTGATGCGCATAATACGTCGCATTTCCCTATCTTTGCGACTTAGGTCTGCTACTGTCCGTATTCTAT
GGGTGTGGTTGGTGAAATCGCCATTTGTATTATCATGCCAATCCAGAGTGAAGAACCTCGTTTAGTTCTTATCGTTCAAGACTGTCACCATTTTTCTTGCCATTTTCACAT
CATTGCTGTTCTGATTCTGCAAATTTTCGAATCATCATTTAAATATCTTCCTTTCGCGAGATCGTCCCAAGATAACATCTTAGAAAACATTAATAACTATGTGATCATCTA
CCAAATACCAGCAGACTTGCGACCGATGTGATCATTCACCAAAGATCAGCTACCAATCCGCATCCCGTATTGTTGATTGAGATAAAGCGGTTGTTCTTTTAAATTACTTTT
TGATGCTATAAGACGTTAGAAAGTACACTCTAGCATTTTGAATAAGCTACTAGCTGTCATAGTCTGCTCTTTTAAGCTCGACTGGCCTTGCGACGTATTTACTACCAGGCA
GTTTGTATATGCTCATGGTGGGTCCGAATTAATGGCTTCCTGTATGTGGCATGCACGTGCTATCCCTGAACCGCTAAGTTCCGCAGGAATAACCGCACGAACTGTAATGAA
CAAGCACAGCCCTGCCTGAATGACTAATGTACACCTCGTTAAATGTACTACCGTGACGATCATAAAGAATGCAATTGCGGCTGCCATTCGTTTAAATTCTGAACCAACTGG
ATCATTAAACGCCAGAGATTATGACTACAAAGGAAGCTAAAAAAGGGGCGTGAAATTTACATTAACTTACACATATTTTAATTGTGGATGAAAGGCGCCTTTTATACATTG
CGCCCACACGCACCCGGTAGGCCAGATTAAATTTCATTTAGGACAAGCTTACTGATACTCCCGTAATTATTATTCACCGAAGCCAAAAGGAAGATAATTGCAGGTTTGAGT
TTCCTAGCGTATTATACGCCTATCTTTAAAATAAGAGGTCAAGTTCACGGTTTCCGACCTTCAAGATTTTGAATATTGTGCGGCGGTAACCTTTAAGTTTTAACGTTTGAT
GGATTGGACACAGCATTGCGTATTGCTTAACATACGAATTAGGGATGAGAATGCAGTTAATCAAATGTGATGCAAATTTGTGTTTCCTATTACGTATTTATTATTTTATAT
AAATTTAAAATAATAAATACGTAACAAGCATATGGTTCCAATTTAGTAGTATTCATTGATCATTTATTTGAACGTAATGTGTGGAACTTATTTCTACTTGTGAATTCCTAA
CTGTGAGCATGTTTTTTTCGTTCACCACGGGCCAGCCACTCAAACACGGAACCGTTGTTGCTCAGAAAAATAAATAGACTGATAAAGTGGAAACAACACTTTTTGTTTCTG
CAGGTTAAGAAAAAAGCTCGGGTGACATTAGATACCAGTTAGATCATAGAGATAATAGCGAAGTTTTCTCAATAACATCATCCGCCATCCGTCTGCCGCGTGATAAACAAA
AAACGAACAAACATTCGTGACAACTAATAATACGCCCCAATGAAGACACGCCGTAAAAAAAGATGATACTGATTGGAACAGGATACAAAAAAAGCAGAAGAAGAAGACTAA
TAGTAGCTAAATTTTGATACGGAATTCGTGTTGACGACGTAAGGAAGTATATAAAGTAATAAATAAAAAATAATAAGAATACATAAAAATAAATCATTTGTTGTACATATGT
ACATAAACCTATAGCGGGGGCTTCGTGTATCACGAAATTGCTATTTTGGCTAAGCAGAGGCTCAATGTTTTACTATAAAATTTTTCCATCAACAGCATGAAATCTTGAGCG
TTTCTCTAATTACATGTAAGATTAACAGCTACTGCACCCCATCGAATTAATGTATTGTCAGCTTCTATAAACGGAAGAGTTAGTTATTAACTCTTCCGTTCCTATAGGGCA
AACATTCTACCAAAGAAAGGAAAGCTAAACGTTTAAGCTGCATTATATTACGTCGTTTAATGAGAAATTTGAGTTAGTGAACTTATACTTTTCAAATTTTTTTCACAAGCT
TGTTTGATTTTCAGTCAAAATATGTACTATTCTTATAGATGGCCTTCTACCATTTAGATAGGAACTTATTTATTCTACTGTTGTTCAAATCTGAAGTATGGGAGCTTATAA
AGTGTCGAATAGCCAAAGAATTTGAATAATTTCTAATAAGAAGATGCGTGGATTTCGAAAAAATCTGGTGAATATGGTGAGTATCAGGAGGTGCTAACTGTTAAGTCCTGT
ATTTTTACTCCTTTGATTATAATCTAAGGCCGAAGGCTGTTTTACGAATAGAGTTGTACCTTTGTTCAGTAATCTTTGTTGCTTAATATGGGGTTGTCCGTCATCATCTTC
TCCAATTATTAAAGATGCACTTCAGCGAAAATCGACGAACCAAAACTCATGAAACTATATAGTTGATAACACCTGTGCTACATGATCAAATAGGCACCACTAGTTGCACCG
TTGCAGCTGTAAAAATATTTTAACGTAATGACCCGCCGGTTCGACATACTGGTATAACAGCCTGGAATTATTGTAAAAGATAAATAATTTTTCACCATGGTCACAATTTGA
TACAAGAACGGGTCACTGTTATAACAGGAGAGCAAATAGACGCCTGCTTAAAATGTATAGCACTCGTTTGTCATTTCTTTTTACTTTTTTTTAATTAGACAGAGATAAATA
AGTATAGTATTTTTA
```

B) Multiple sequence:

```
D. melanogaster  MEHHNSHLLPGGSEKMYYIAHQQPMLRNEDDNYQEGYFIRPDPASLIYNTTALPADDEGS
A. mellifera     ---------------------------MNNETNGSLDG------PTGYKLNLTN-PDD----
I. scapularis    ----------------------------MDSTNGPSA-------PPTATSNWTSQPAS----
T. urticae       ------------------------------------------------------------
V. destructor    ---------------------------MSAAAGGP---------PSGGVNSITGLSSA----
V.jacobsoni      ---------------------------MSAAAGGP---------PSGGVNSITGLSSA----

D. melanogaster  NYGYGSTTTLSGLQFETYNITVMMNFSCDDYDLLSEDMWSSAYFKII--VYMLYIPIFIF
A. mellifera     ----------------------------CNYFE--SQSIFSTNWFRGI--IYFVYSTVFVV
I. scapularis    ----------------------TESAACDL------PPPVPEGMQAL--MYIMYIAVSVA
T. urticae       ------------------------------------MKF-PMLAEVL--MYFMYSSICVI
V. destructor    ----------------------TTARINCS-----LDGPWEPAQVPALVVMSIMYGLVTLF
V.jacobsoni      ----------------------TTARINCS-----LDGPWEPAQVPALVVMSIMYGLVTLF

D. melanogaster  ALIGNGTVCYIVYSTPRMRTVTNYFIASLAIGDILMSFFCVPSSFISLFILNYWPFGLAL
A. mellifera     ALTGNGLVCYVVHSSPRMKTVTNFFIVNLALGDILIALFCVPTSSISTLILQYWPFGPEL
I. scapularis    AIGGNGIVCYIVLAYQRMRTVTNMFIMNLAIGDILMASLCIPFTFVSNLLLGYWPFGGVM
T. urticae       AITGNAIVCFIVIAYRRMQSVTNYFIVNLALSDILMASLCIPFTFVSNLILQHWPFGAFM
V. destructor    ALAGNGLVCYIVLAYPRMRTVTNMFIVNLAVGDILMAVFCIPFSFVSNFITEHWVFGYFM
V.jacobsoni      ALAGNGLVCYIVLAYPRMRTVTNMFIVNLAVGDILMAVFCIPFSFVSNFITEHWVFGYFM

D. melanogaster  CHFVNYSQAVSVLVSAYTLVAISIDRYIAIMWPLKPRITKRYATFIIAGVWFLALATALP
A. mellifera     CPTVIYLQAVSVLVSAYTLVAISIDRYIAIMWPLKPRLSKRQAQLLILAVWMLAMLISLP
I. scapularis    CVVVTYAQCVTVFISAYTLIAISVDRYTAIVYPLRPRMTKLRSKIIIGVVWLVALVTPLP
T. urticae       CVLVSYSQAVSVFISAYTLIALSLDRYIAILYPLRPKMTRFQAKLIIACVWIVALITPLP
V. destructor    CIAVGYCQAISVMVSAYTLIAISMDRYLAIMYPLKPRLNSLQARLIIGIVWAVALVTPLP
V.jacobsoni      CIAVGYCQAISVMVSAYTLIAISMDRYLAIMYPLKPRLNSLQARLIIGIVWAVALVTPLP
```

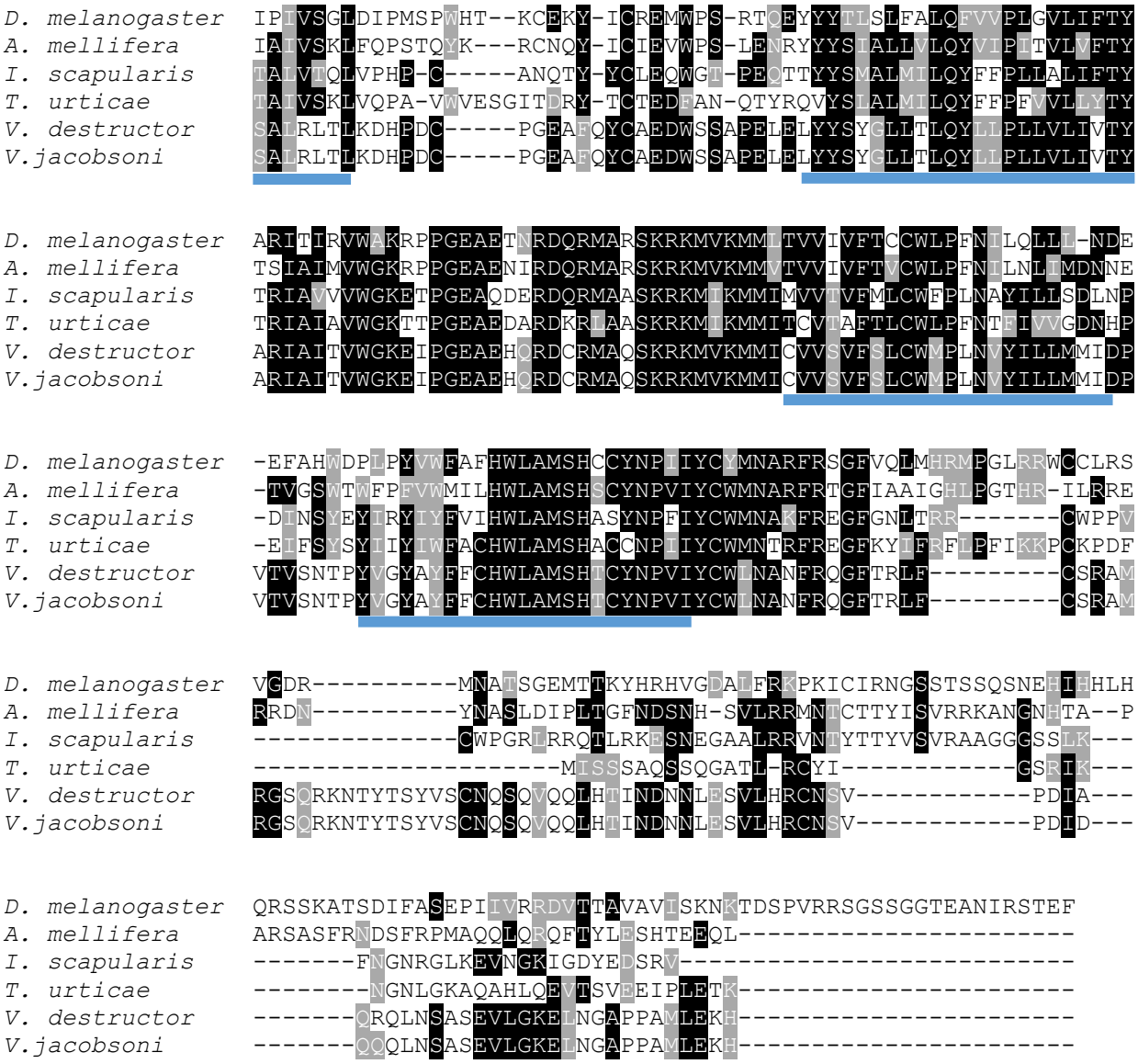


Figure 3.11 A) Gene structure for RYamide receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of RYamide receptor from various orthologous species across Phylum Arthropoda.

CCHamide receptor

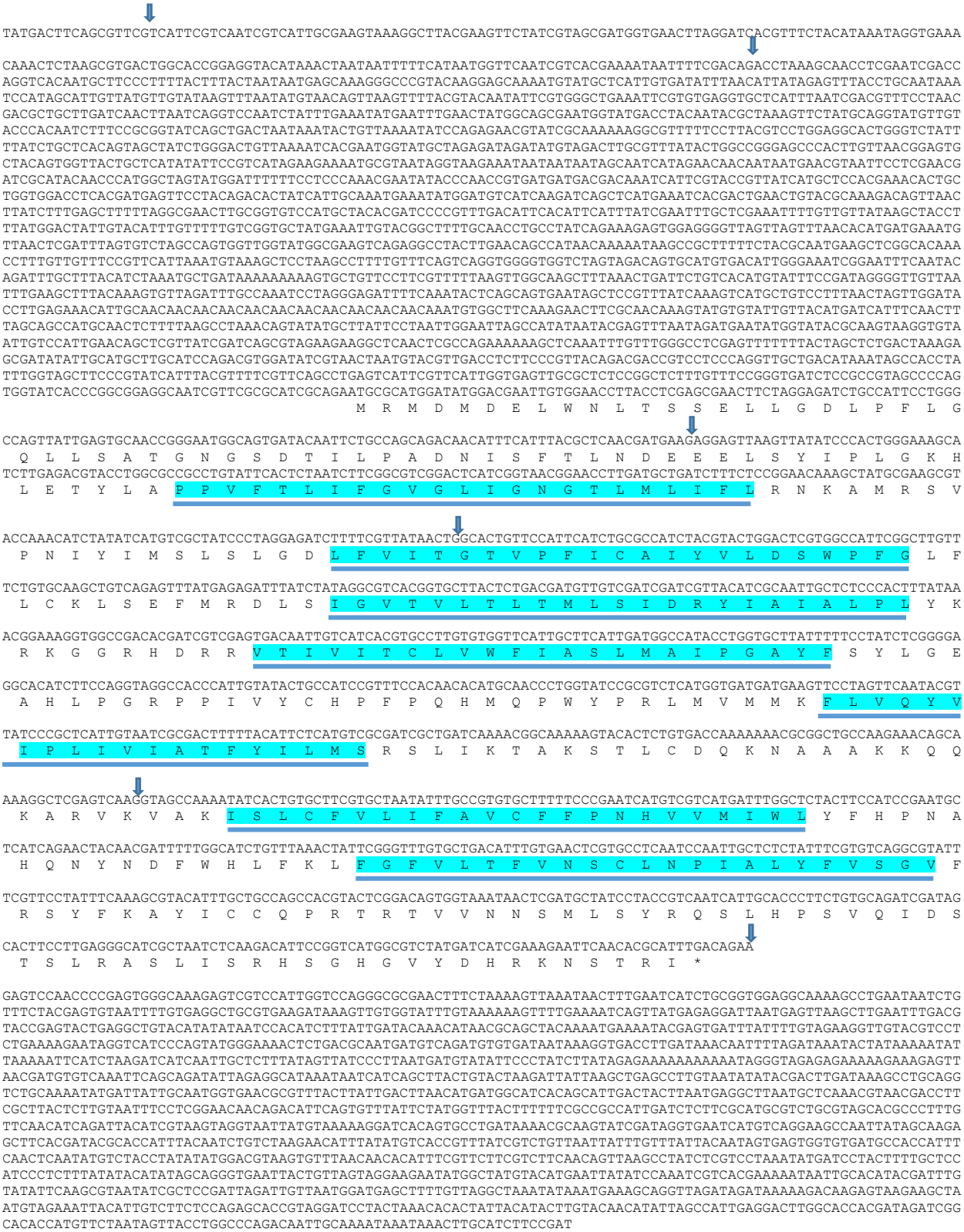
A) Gene structure:

V. destructor

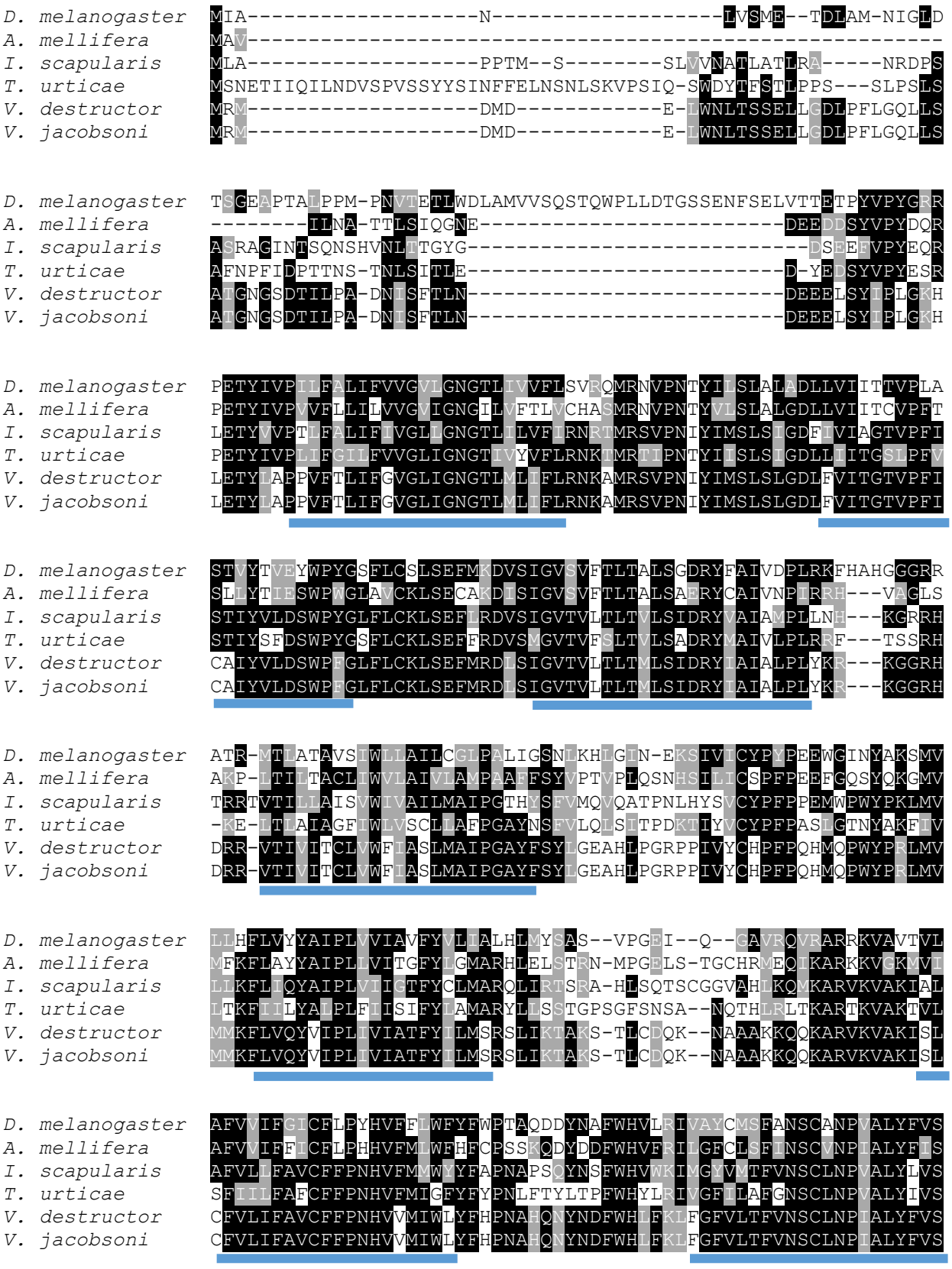
```
ACGACCGTCCTCCCAGGTTGCTGACATAAATAGCCACCTATTTGGTAGCTTTCCGTATCATTTACGTTTTCGTTCAGCCTGAGCCATTCGTTCATTGGTGAGTTGCGCTCT
CCGGCTCTTTGTTTCCGGGTGATCTCCGCCGTAGCCCCAGTGGTATCACCCGGCGGAGGCAATCGTTCGCGCATCGCAGAATGCGCATGGATATGGACGAATTGTGGAACC
                                                                                    M  R  M  D  M  D  E  L  W  N
TTACCTCGAGCGAACTTCTAGGAGATCTGCCATTCCTGGGCCAGTTATTGAGTGCAACCGGGAATGGCAGTGATACAATTCTGCCAGCAGACAACATTTCATTTACGCTCA
L  T  S  S  E  L  L  G  D  L  P  F  L  G  Q  L  L  S  A  T  G  N  G  S  D  T  I  L  P  A  D  N  I  S  F  T  L

ACGATGAAGAGGAGTTAAGTTATATCCCACTGGGAAAGCATCTTGAGACGTACCTGGCGCCGCCTGTATTCACTCTAATCTTCGGCGTCGGACTCATCGGTAACGGAACCT
N  D  E  E  E  L  S  Y  I  P  L  G  K  H  L  E  T  Y  L  A  P  P  V  F  T  L  I  F  G  V  G  L  I  G  N  G  T

TGATGCTGATCTTTCTCCGGAACAAAGCTATGCGAAGCGTACCAAACATCTATATCATGTCGCTATCCCTAGGAGATCTTTTCGTTATAACTGGCACTGTTCCATTCATCT
L  M  L  I  F  L  R  N  K  A  M  R  S  V  P  N  I  Y  I  M  S  L  S  L  G  D  L  F  V  I  T  G  T  V  P  F  I

GCGCCATCTACGTACTGGACTCGTGGCCATTCGGCTTGTTTCTGTGCAAGCTGTCAGAGTTTATGAGAGATTTATCTATAGGCGTCACAGTGCTTACTCTGACGATGTTGT
C  A  I  Y  V  L  D  S  W  P  F  G  L  F  L  C  K  L  S  E  F  M  R  D  L  S  I  G  V  T  V  L  T  L  T  M  L

CGATCGATCGTTACATCGCAATTGCTCTCCCACTTTATAAACGGAAAGGTGGCCGACACGATCGTCGAGTGACAATTGTCATCACGTGCCTTGTGTGGTTCATTGCTTCAT
S  I  D  R  Y  I  A  I  A  L  P  L  Y  K  R  K  G  G  R  H  D  R  R  V  T  I  V  I  T  C  L  V  W  F  I  A  S

TGATGGCCATACCTGGTGCTTATTTTTCCTATCTCGGGGAGGCACATCTTCCAGGTAGGCCACCCATTGTATACTGCCATCCGTTTCCACAACACATGCAACCCTGGTATC
L  M  A  I  P  G  A  Y  F  S  Y  L  G  E  A  H  L  P  G  R  P  P  I  V  Y  C  H  P  F  P  Q  H  M  Q  P  W  Y

CGCGTCTCATGGTGATGATGAAGTTCCTAGTTCAATACGTTATCCCGCTCATTGTAATCGCGACTTTTTACATTCTCATGTCGCGATCGCTGATCAAAACGGCAAAAAGTA
P  R  L  M  V  M  M  K  F  L  V  Q  Y  V  I  P  L  I  V  I  A  T  F  Y  I  L  M  S  R  S  L  I  K  T  A  K  S

CACTCTGTGACCAAAAAAACGCGGCTGCCAAGAAACAGCAAAAGGCTCGAGTCAAGGTAGCCAAAATATCACTGTGCTTCGTGCTAATATTTGCCGTGTGCTTTTTCCCGA
T  L  C  D  Q  K  N  A  A  A  K  K  Q  Q  K  A  R  V  K  V  A  K  I  S  L  C  F  V  L  I  F  A  V  C  F  F  P

ATCATGTCGTCATGATTTGGCTCTACTTCCATCCGAATGCTCATCAGAACTACAACGATTTTTGGCATCTGTTTAAACTATTCGGGTTTGTGCTGACATTTGTGAACTCGT
N  H  V  V  M  I  W  L  Y  F  H  P  N  A  H  Q  N  Y  N  D  F  W  H  L  F  K  L  F  G  F  V  L  T  F  V  N  S

GCCTCAATCCAATTGCTCTCTATTTCGTGTCAGGCGTATTTCGTTCCTATTTCAAAGCGTACATTTGCTGCCAGCCACGTACTCGGACAGTGGTAAATAACTCGATGCTAT
C  L  N  P  I  A  L  Y  F  V  S  G  V  F  R  S  Y  F  K  A  Y  I  C  C  Q  P  R  T  R  T  V  V  N  N  S  M  L

CCTACCGTCAATCATTGCACCCTTCTGTGCAGATCGATAGCACTTCCTTGAGGGCATCGCTAATCTCAAGACATTCCGGTCATGGCGTCTATGATCATCGAAAGAATTCAA
S  Y  R  Q  S  L  H  P  S  V  Q  I  D  S  T  S  L  R  A  S  L  I  S  R  H  S  G  H  G  V  Y  D  H  R  K  N  S
CACGCATTTGACAGAA
T  R  I  *

CCTGGTAATCACCTTTGCCAGCTCTGGGAATTGAACCAGCGCGGCTGCACGCATCATATTCTTTTCATGGTAAACATAACGACGTAGTTTAAAAGACGCGTATTAGTCACG
CGTCTTTCTCGCTGTAACAGCAACCTATTACATGTCGCAAACATGCGGAGCTAACCACACTAGTTCCACATACAGTAAAAAAGAAAAAAACTATTTGTCCCCGTTCCCTCA
GATGCTGAAGTTTCAAATCAAGGTAATTTCTCAGTTTTTCGGTGATCTTTCATAAATTATCAGCTTTGTCATTTGAATTTAGTAATCCACTATGCTCAGCTACTTACCACT
ACTGGATAACTACAATATATGATCAAATTAAAGAAGTATTCGCAATGTTCAGTCAATATCACATAAATATATGTGGTCTTTGACTAGGGTCTCATTAGTGTTAGTAGGGTT
CCCCAGTTATTTTTATATAACCCCATAATGATTTTTAATGTCTAGTTTCAATATATTGTGCAGTTGAAAAATCTTGCGGAAATGTCCCGCACTTCCTGCAGATTTCAATCG
ACAGGAACAAAACTACATATATATATATATATATAAGCATCATATTAAAAAAATCACTCACAGATGATCTCGTTACTTACTATTCATTTATTATTCTTGTATTGTGATATC
GGTTCACTTGATAGTGAGCTTTCCCTGGACTCAAATTCGATGGGCTATTTTTAGAATCGTTCTTATTTTAGATGTCCTCTATCTTTTTCAATATGTTTCTATCTGTCTATG
TCATAATCATACGAATAAAAAGCGCCACTTAATAAGCATCTATCTGGAATAGCACACCGTATGTCACCAAAATAAAACCATTCATTACTTACTGCGTTTACTTCACAAACA
ATATTGCCCTGTAGGGCTTAGTGTCATCTCAAATAAAAAAAAATGCTTGTGAGCACTAGAATCTTGTTTAATAGAGAACACAAAAACCTATTTTCAAATGTTTAAAATGCT
TGATTCTCTTTTCTCGCTAAAACGAATCTGGGGGCCGTCTTTTGAAGCATAACACAGCAGATTGTCTTGATTCACCGCTGCGGTGAGAGCCTTCCCTGTACTCAATCACGG
AAAGCGGCCCGCATCTTCACGATTACTTCTTGTCTACACAGATAACAAAATGACTCTGAAAATAACTACAGTTTCAATAAGCTGAAGGCTTGATATTAAAAAGAAAAAAAA
ACTGTGTGTTATGTGACACATATTTGCTGTATTTGATTTGCTTGGAGCAACTTGGATGAAAACAAAGTCGCTTTTTTTAGCTACCTTTTGTCTCAACATCTCATAGAATTA
ATGCGTTATTGGTAATCGCCTTCGCTGAAGGCATGAGCGATATCACCGACACATGTTGCAATTGCAAGTCTATTGTACACCCGAAGCTTGTGCCTGATTCTGCGAATCGAG
GAAACACGTTACACGAAATGATGGCCATCATAAGAAAAACATCTAAAATTCGCCAAGTTTATCTATCTCCATAGTGATCTTGACTTAATTCGCACAAATCGTTTTGCTGTT
ATTAATAGTAGTTTTCATTTTATTGTAAGTAAATTTAACAAAAAAAAAAAGACAAGCTAGTTACTAGCATTAAATTTAATAATGCATTTCGATTCTATATATTTTTAAATG
ATTGTACATAATGGTCTGATACGCTGGTGTGGTAGAGATAATGGTCTGGCCGGGACATGACAGGCGTTTGCACGGAATTCAGCAAACTTCCTCGGACACTTTATATTTAAG
TTAGACTAGGACCCGTTTTGCACTTTTCAGAACGAGGTAGAATGTCAAATATATTATCAGCTACGGTACATATTACTAGAGCACCAGTGTTGACTGCATTACCCGACTCGA
AAAATCGTCGCGAATGCAACGTAATTTATATTCAATGTCAGTCTGTATAGCAGCTACAGCGTTTACGTATCCGAGAGAACTGAGTATGGTGTAACGCTTTGCAGGTCTGAT
CTGGACAAGCTCAGCGACATTTACGCTCACAGCCTTTCCATAAGGCAAGGACATCGCCCCATTTGTCTCAAATACAGCACCTGTTGTAGCAGCAGCAGAAAGATTCACATG
GAAAAAATAAGTAACTGTTCTGCGTAGACCTAGAGGCACACGATCGACATTGCGAGGATGCAGAAGGGACGATCGAATGACTCTTGGCAGCAAAGTAAATTTTTCTGACAG
AACCTCTGCATATAGCAGGCGATCGACGGGCCCACAATGACACACATATTTTTCGGGGCCAACAGGGAACATTATGGCAAAATACTCATTCTCAGCAGCAGAAGCAACAAC
TGGGTTATTGTCAATCGTTTGGATGCCTACCGTTTGGATAATTCATCGCGCTGAAGCTTTCTCAGACGCAACAAAGACATGTAGCAATGAAGTGGTCTCTAACGAATGTGG
ACATGGGCAAAATGGAACTCGAGCAGAACAAACAGGTACAAGAACACGTTCTTTTATGCACTGATGTTTATGCCGAATCGAGTGATGCCTACCCTTGCAAAGAGAACCCAC
TTTTTATTCAAACATAAAAATTAGAAAAGACTGCGAAGGGTGGTACATTCGATCCTACATACACATAATAACAAAAAAAAATAGCTTTCAAAGTCTACTATTAACATGACT
GGTGTACATACTAAGCATAAGTAAACGAAAAATGTGTGTCCACTTTAATAAAACCCGAGGTTTAACGGTTCGG
```

V. jacobsoni

```
ACGGTCTGCTTCTACAGGGCATGTTGTCTTGATGGTGTACGGGGGGTGAGTGACAAGGTCGCACTCAAAAGCTAACTTCCCTACCTCCCAGTGACCCCTAAAGAGCCATCC
GAAAAAATCAATCCAGGAGGGAGGGCCCTTCACTCGCAAGTGCGTGAAGACGCGCACCGTTCGAATCATAATGTACGGAGTCCAGCAGCAGTATTCCATATTGTTTCATGC
```


TATGACTTCAGCGTTCGTCATTCGTCAATCGTCATTGCGAAGTAAAGGCTTACGAAGTTCTATCGTAGCGATGGTGAACTTAGGATCACGTTTCTACATAAATAGGTGAAA
CAAACTCTAAGCGTGACTGGCACCGGAGGTACATAAACTAATAATTTTTCATAATGGTTCAATCGTCACGAAAATAATTTTCGACAGACCTAAAGCAACCTCGAATCGACC
AGGTCACAATGCTTCCCTTTTACTTTACTAATAATGAGCAAAGGGCCCGTACAAGGAGCAAAATGTATGCTCATTGTGATATTTAACATTATAGAGTTTACCTGCAATAAA
TCCATAGCATTGTTATGTTGTATAAGTTTAATATGTAACAGTTAAGTTTTACGTACAATATTCGTGGGCTGAAATTCGTGTGAGGTGCTCATTTAATCGACGTTTCCTAAC
GACGCTGCTTGATCAACTTAATCAGGTCCAATCTATTTGAAATATGAATTTGAACTATGGCAGCGAATGGTATGACCTACAATACGCTAAAGTTCTATGCAGGTATGTTGT
ACCCACAATCTTTCCGCGGTATCAGCTGACTAATAAATACTGTTAAAATATCCAGAGAACGTATCGCAAAAAAGGCGTTTTTCCTTACGTCCTGGAGGCACTGGGTCTATT
TTATCTGCTCACAGTAGCTATCTGGGACTGTTAAAATCACGAATGGTATGCTAGAGATAGATATGTAGACTTGCGTTTATACTGGCCGGGAGCCCACTTGTTAACGGAGTG
CTACAGTGGTTACTGCTCATATATTCCGTCATAGAAGAAAATGCGTAATAGGTAAGAAATAATAATAATAGCAATCATAGAACAACAATAATGAACGTAATTCCTCGAACG
ATCGCATACAACCCATGGCTAGTATGGATTTTTTCCTCCCAAACGAATATACCCAACCGTGATGATGACGACAAATCATTCGTACCGTTATCATGCTCCACGAAACACTGC
TGGTGGACCTCACGATGAGTTCCTACAGACACTATCATTGCAAATGAAATATGGATGTCATCAAGATCAGCTCATGAAATCACGACTGAACTGTACGCAAAGACAGTTAAC
TTATCTTTGAGCTTTTTAGGCGAACTTGCGGTGTCCATGCTACACGATCCCCGTTTGACATTCACATTCATTTATCGAATTTGCTCGAAATTTTGTTGTTATAAGCTACCT
TTATGGACTATTGTACATTTGTTTTTGTCGGTGCTATGAAATTGTACGGCTTTTGCAACCTGCCTATCAGAAAGAGTGGAGGGGGTTAGTTAGTTTAACACATGATGAAATG
TTAACTCGATTTAGTGTCTAGCCAGTGGTTGGTATGGCGAAGTCAGAGGCCTACTTGAACAGCCATAACAAAAATAAGCCGCTTTTTCTACGCAATGAAGCTCGGCACAAA
CCTTTGTTGTTTCCGTTCATTAAATGTAAAGCTCCTAAGCCTTTTGTTTCAGTCAGGTGGGGTGGTCTAGTAGACAGTGCATGTGACATTGGGAAATCGGAATTTCAATAC
AGATTTGCTTTACATCTAAATGCTGATAAAAAAAAAAGTGCTGTTCCTTCGTTTTTAAGTTGGCAAGCTTTAAACTGATTCTGTCACATGTATTTCCGATAGGGGTTGTTAA
TTTGAAGCTTTACAAAGTGTTAGATTTGCCAAATCCTAGGGAGATTTTCAAATACTCAGCAGTGAATAGCTCCGTTTATCAAAGTCATGCTGTCCTTTAACTAGTTGGATA
CCTTGAGAAACATTGCAACAACAACAACAACAACAACAACAACAACAACAAATGTGGCTTCAAAGAACTTCGCAACAAAGTATGTGTATTGTTACATGATCATTTCAACTT
TAGCAGCCATGCAACTCTTTTAAGCCTAAACAGTATATGCTTATTCCTAATTGGAATTAGCCATATAATACGAGTTTAATAGATGAATATGGTATACGCAAGTAAGGTGTA
ATTGTCCATTGAACAGCTCGTTATCGATCAGCGTAGAAGAAGGCTCAACTCGCCAGAAAAAAGCTCAAATTTGTTTGGGCCTCGAGTTTTTTTACTAGCTCTGACTAAAGA
GCGATATATTGCATGCTTGCATCCAGACGTGGATATCGTAACTAATGTACGTTGACCTCTTCCCGTTACAGACGACCGTCCTCCCAGGTTGCTGACATAAATAGCCACCTA
TTTGGTAGCTTCCCGTATCATTTACGTTTTCGTTCAGCCTGAGTCATTCGTTCATTGGTGAGTTGCGCTCTCCGGCTCTTTGTTTCCGGGTGATCTCCGCCGTAGCCCCAG
TGGTATCACCCGGCGGAGGCAATCGTTCGCGCATCGCAGAATGCGCATGGATATGGACGAATTGTGGAACCTTACCTCGAGCGAACTTCTAGGAGATCTGCCATTCCTGGG
M R M D M D E L W N L T S S E L L G D L P F L G
CCAGTTATTGAGTGCAACCGGGAATGGCAGTGATACAATTCTGCCAGCAGACAACATTTCATTTACGCTCAACGATGAAGAGGAGTTAAGTTATATCCCACTGGGAAAGCA
Q L L S A T G N G S D T I L P A D N I S F T L N D E E E L S Y I P L G K H
TCTTGAGACGTACCTGGCGCCGCCTGTATTCACTCTAATCTTCGGCGTCGGACTCATCGGTAACGGAACCTTGATGCTGATCTTTCTCCGGAACAAAGCTATGCGAAGCGT
L E T Y L A P P V F T L I F G V G L I G N G T L M L I F L R N K A M R S V
ACCAAACATCTATATCATGTCGCTATCCCTAGGAGATCTTTTCGTTATAACTGGCACTGTTCCATTCATCTGCGCCATCTACGTACTGGACTCGTGGCCATTCGGCTTGTT
P N I Y Y I M S L S L G D L F V I T G T V P F I C A I Y V L D S W P F G L F
TCTGTGCAAGCTGTCAGAGTTTATGAGAGATTTATCTATAGGCGTCACGGTGCTTACTCTGACGATGTTGTCGATCGATCGTTACATCGCAATTGCTCTCCCACTTTATAA
L C K L S E F M R D L S I G V T V L T L T M L S I D R Y I A I A L P L Y K
ACGGAAAGGTGGCCGACACGATCGTCGAGTGACAATTGTCATCACGTGCCTTGTGTGGTTCATTGCTTCATTGATGGCCATACCTGGTGCTTATTTTTCCTATCTCGGGGA
R K G G R H D R R V T I V I T C L V W F I A S L M A I P G A Y F S Y L G E
GGCACATCTTCCAGGTAGGCCACCCATTGTATACTGCCATCCGTTTCCACAACACATGCAACCCTGGTATCCGCGTCTCATGGTGATGATGAAGTTCCTAGTTCAATACGT
A H L P G R P P P I V Y C H P F P Q H M Q P W Y P R L M V M M K F L V Q Y V
TATCCCGCTCATTGTAATCGCGACTTTTTACATTCTCATGTCGCGATCGCTGATCAAAACGGCAAAAAGTACACTCTGTGACCAAAAAAACGCGGCTGCCAAGAAACAGCA
I P L I V I A T F Y I L M S R S L I K T A K S T L C D Q K N A A A K K Q Q
AAAGGCTCGAGTCAAGGTAGCCAAAATATCACTGTGCTTCGTGCTAATATTTGCCGTGTGCTTTTTCCCGAATCATGTCGTCATGATTTGGCTCTACTTCCATCCGAATGC
K A R V K V A K I S L C F V L I F A V C F F P N H V V M I W L Y F H P N A
TCATCAGAACTACAACGATTTTTGGCATCTGTTTAAACTATTCGGGTTTGTGCTGACATTTGTGAACTCGTGCCTCAATCCAATTGCTCTCTATTTCGTGTCAGGCGTATT
H Q N Y N D F W H L F K L F G F V L T F V N S C L N P I A L Y F V S G V F
TCGTTCCTATTTCAAAGCGTACATTTGCTGCCAGCCACGTACTCGGACAGTGGTAAATAACTCGATGCTATCCTACCGTCAATCATTGCACCCTTCTGTGCAGATCGATAG
R S Y F K A Y I C C Q P R T R T V V N N S M L S Y R Q S L H P S V Q I D S
CACTTCCTTGAGGGCATCGCTAATCTCAAGACATTCCGGTCATGGCGTCTATGATCATCGAAAGAATTCAACACGCATTTGACAGAA
T S L R A S L I S R H S G H G V Y D H R K N S T R I *
GAGTCCAACCCCGAGTGGGCAAAGAGTCGTCCATTGGTCCAGGGCGCGAACTTTCTAAAAGTTAAATAACTTTGAATCATCTGCGGTGGAGGCAAAAGCCTGAATAATCTG
TTTCTACGAGTGTAATTTTGTGAGGCTGCGTGAAGATAAAGTTGTGGTATTTGTAAAAAAGTTTTGAAAATCAGTTATGAGAGGATTAATGAGTTAAGCTTGAATTTGACG
TACCGAGTACTGAGGCTGTACATATATAATCCACATCTTTATTGATACAAACATAACGCAGCTACAAAATGAAAATACGAGTGATTTATTTTGTAGAAGGTTGTACGTCCT
CTGAAAAGAATAGGTCATCCCAGTATGGGAAAACTCTGACGCAATGATGTCAGATGTGTGATAATAAAGGTGACCTTGATAAACAATTTTAGATAAATACTATAAAAATAT
TAAAAATTCATCTAAGATCATCAATTGCTCTTTATAGTTATCCCTTAATGATGTATATTCCCTATCTTATAGAGAAAAAAAAAAAATAGGGTAGAGAGAAAAAGAAAGAGTT
AACGATGTGTCAAATTCAGCAGATATTAGAGGCATAAATAATCATCAGCTTACTGTACTAAGATTATTAAGCTGAGCCTTGTAATATATACGACTTGATAAAGCCTGCAGG
TCTGCAAAATATGATTATTGCAATGGTGAACGCGTTTACTTATTGACTTAACATGATGGCATCACAGCATTGACTACTTAATGAGGCTTAATGCTCAAACGTAACGACCTT
CGCTTACTCTTGTAATTTCCTCGGAACAACAGACATTCAGTGTTTATTCTATGGTTTACTTTTTTCGCCGCCATTGATCTCTTCGCATGCGTCTGCGTAGCACGCCCTTTG
TTCAACATCAGATTACATCGTAAGTAGGTAATTATGTAAAAAGGATCACAGTGCCTGATAAAACGCAAGTATCGATAGGTGAATCATGTCAGGAAGCCAATTATAGCAAGA
GCTTCACGATACGCACCATTTACAATCTGTCTAAGAACATTTATATGTCACCGTTTATCGTCTGTTAATTATTTGTTTATTACAATAGTGAGTGGTGTGATGCCACCATTT
CAACTCAATATGTCTACCTATATATGGACGTAAGTGTTTAACAACACATTTCGTTCTTCGTCTTCAACAGTTAAGCCTATCTCGTCCTAAATATGATCCTACTTTTGCTCC
ATCCCTCTTTATATACATATAGCAGGGTGAATTACTGTTAGTAGGAAGAATATGGCTATGTACATGAATTATATCCAAATCGTCACGAAAAATAATTGCACATACGATTTG
TATATTCAAGCGTAATATCGCTCCGATTAGATTGTTAATGGATGAGCTTTTGTTAGGCTAAATATAAATGAAAGCAGGTTAGATAGATAAAAAGACAAGAGTAAGAAGCTA
ATGTAGAAATTACATTGTCTTCTCCAGAGCACCGTAGGATCCTACTAAACACACTATTACATACTTGTACAACATATTAGCCATTGAGGACTTGGCACCACGATAGATCGG
CACACCATGTTCTAATAGTTACCTGGCCCAGACAATTGCAAAATAAATAAACTTGCATCTTCCGAT

B) Multiple sequence alignment :



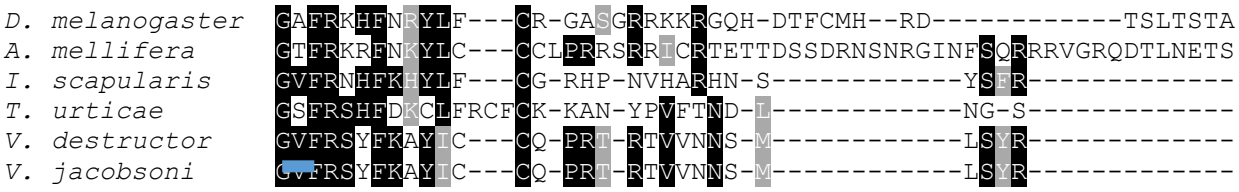
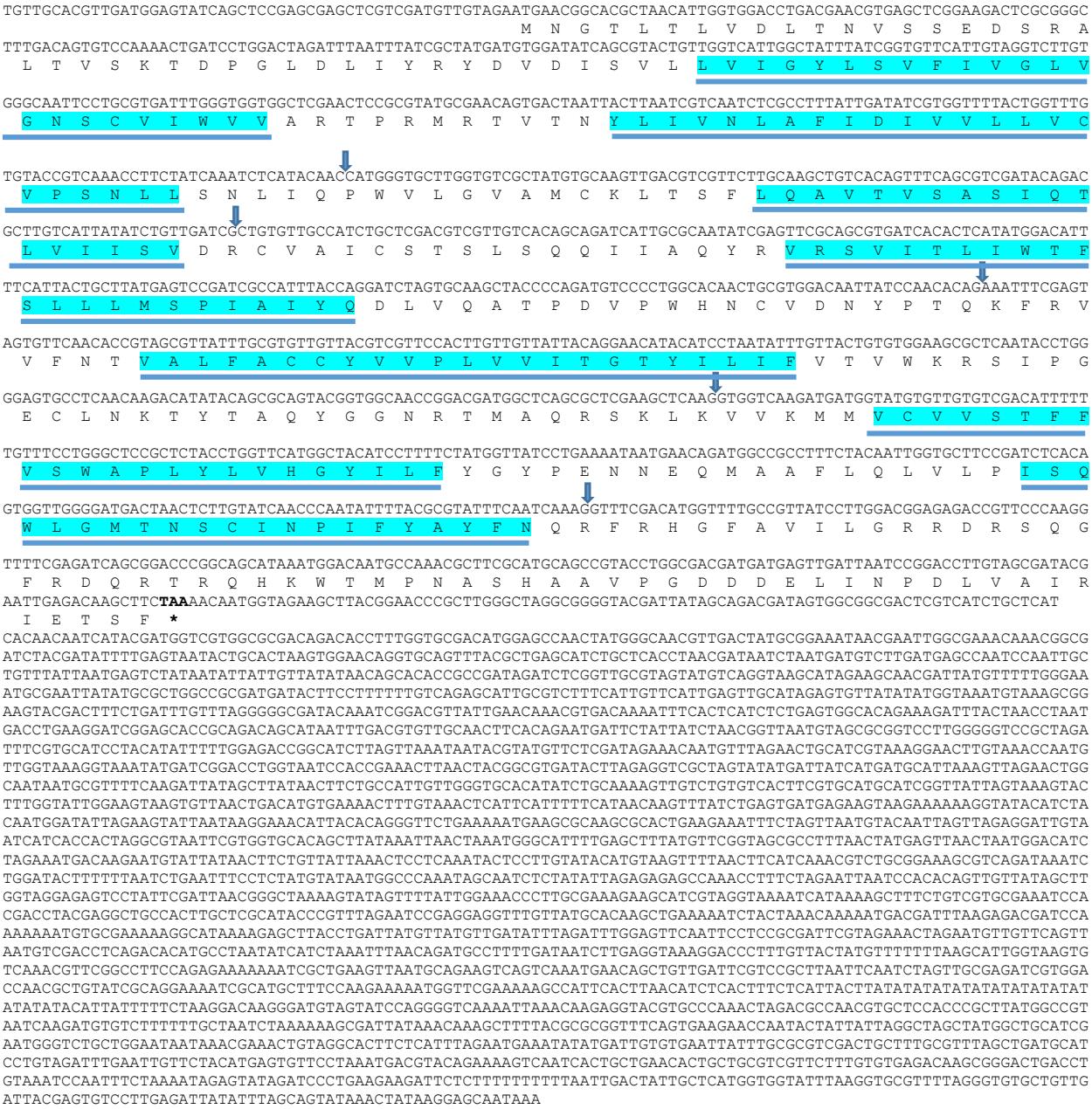


Figure 3.12 A) Gene structure for CCHamide receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of CCHamide receptor from various orthologous species across Phylum Arthropoda.

SIFamide receptor

A) *V. destructor*

Gene structure:

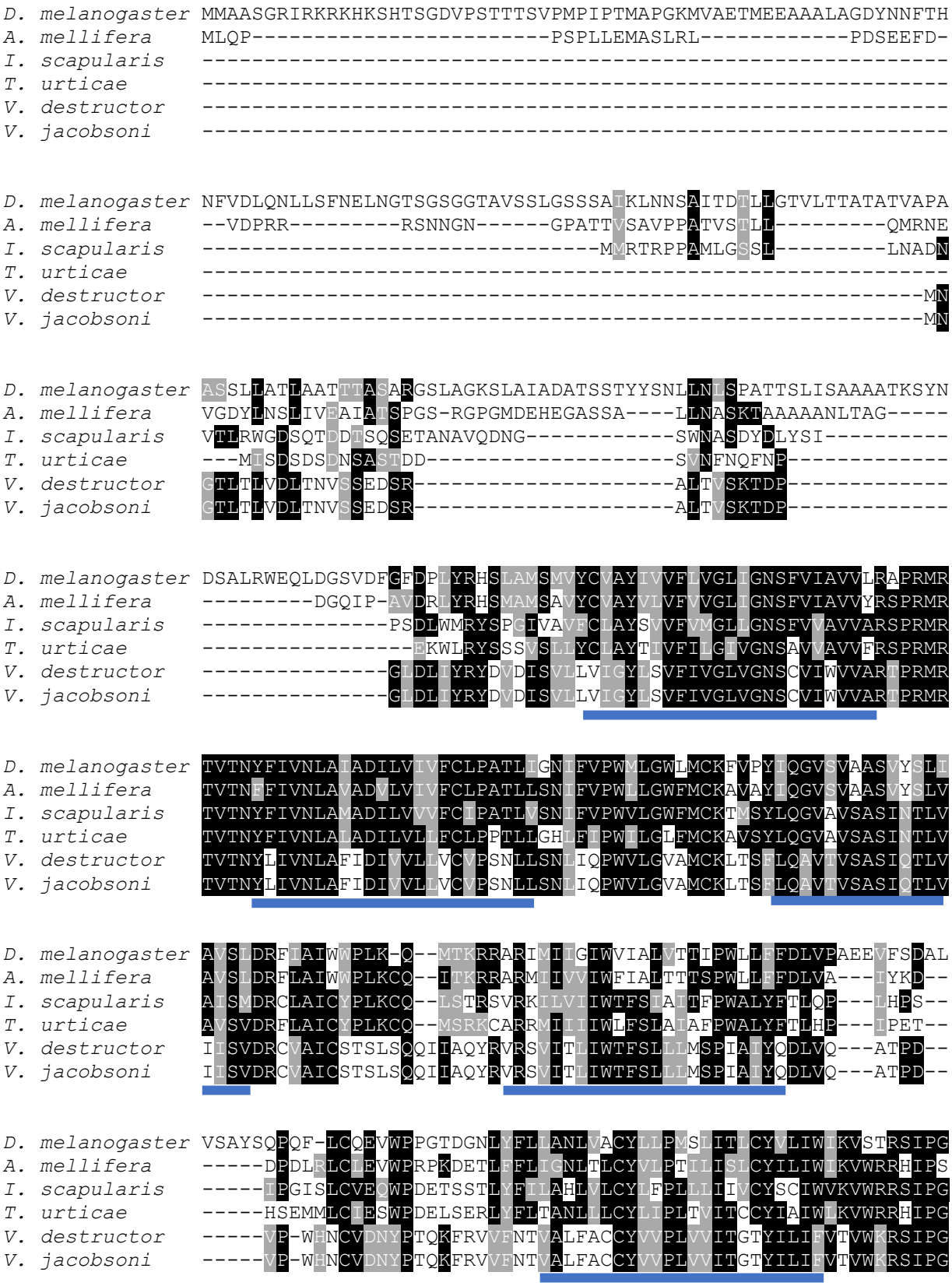


V. jacobsoni

Gene structure:


GCGAACAGTGACTAATTACTTAATCGTCAATCTCGCCTTTATTGATATCGTGGTTTTACTGGTTTGTGTACCGTCAAACCTTCTATCAAATCTCATACAACCATGGGTGCT
R T V T N Y L I V N L A F I D I V V L L V C V P S N L L S N L I Q P W V L
TGGTGTCGCTATGTGCAAGTTGACGTCGTTCTTGCAAGCTGTCACAGTTTCAGCGTCGATACAGACGCTTGTCATTATATCTGTTGATCGCTGTGTTGCCATCTGCTCGAC
G V A M C K L T S F L Q A V T V S A S I Q T L V I I S V D R C V A I C S T
GTCGTTGTCACAGCAGATCATTGCGCAATATCGAGTTCGCAGCGTGATCACACTCATATGGACATTTTCATTACTGCTTATGAGTCCGATCGCCATTTACCAGGATCTAGT
S L S Q Q I I A Q Y R V R S V I T L I W T F S L L L M S P I A I Y Q D L V
GCAAGCTACCCCAGATGTCCCCTGGCACAACTGCGTGGACAATTATCCAACACAGAAATTTCGAGTAGTGTTCAACACCGTAGCGTTATTTGCGTGTTGTTACGTCGTTCC
Q A T P D V P W H N C V D N Y P T Q K F R V V F N T V A L F A C C Y V V P
ACTTGTTGTTATTACAGGAACATACATCCTAATATTTGTTACTGTGTGGAAGCGCTCAATACCTGGGGAGTGCCTCAACAAGACATATACAGCGCAGTACGGTGGCAACCG
L V V I T G T Y I L I F V T V W K R S I P G E C L N K T Y T A Q Y G G N R
GACGATGGCTCAGCGCTCGAAGCTCAAGGTGGTCAAGATGATGGTATGTGTTGTGTCGACATTTTTTGTTTCCTGGGCTCCGCTCTACCTGGTTCATGGCTACATCCTTTT
T M A Q R S K L K V V K M M V C V V S T F F V S W A P L Y L V H G Y I L F
CTATGGTTATCCTGAAAATAATGAACAGATGGCCGCCTTTCTACAATTGGTGCTTCCGATCTCACAGTGGTTGGGGATGACTAACTCTTGTATCAACCCAATATTTTACGC
Y G Y P E N N E Q M A A F L Q L V L P I S Q W L G M T N S C I N P I F Y A
GTATTTCAATCAAAGGTTTCGACATGGTTTTGCCGTTATCCTTGGACGGAGAGACCGTTCCCAAGGTTTTCGAGATCAGCGGACCCGGCAGCATAAATGGACAATGCCAAA
Y F N Q R F R H G F A V I L G R R D R S Q G F R D Q R T R Q H K W T M P N
CGCTTCGCATGCAGCCGTACCTGGCGACGATGATGAGTTGATTAATCCGGACCTTGTAGCGATACGAATTGAGACAAGCTTCTAAACAATGGTAGAAGCTTACGGAACCC
A S H A A V P G D D D E L I N P D L V A I R I E T S F *
GCTTGGGCTAGGCGGGGTACGATTATAGCAGACGATAGTGGCGGCGACTCGTCATCTGCTCATCACAACAATCATACGATGGTCGTGGCGCGACAGACACCTTTGGTGCGA
CATGGAGCCAACTGTGGGCAACGTTGACTATGCCGAAATAACGAATTGGCGAAACAAACGGCGATCTACGATATTTTGAGTAATACTGCACTAAGTGGAACAGGTGCAGTT
TACGCTGAGCATCTGCTCACCTAACGATAATCTAATGATGTCTTGATGAGCCAATCCAATTGCTGTTTATTAATGAGTCTATAATATTATTGTTATATAACAGCACACCGC
CGATAGATCTCGGTTGCGTAGTATGTCAGGTAAGCATAGAAGCAACGATTATGTTTTTGGGAAATGCGAATTATATGCGCTGGCCGCGATGATACTTCTTTTTTTGTCAGA
GCATTGCGTCTTTCATTGTTCATTGAGTTGCATAGAGTGTTATATATGGTAAATGTAAAGCGGAAGTACGACTTTCTGATTTGTTTAGGGGGCGATACAAATCGGACGTTA
TTGAACAAACGTGACAAAATTTCACTCATCTCTGAGTGGCACAGAAAGATTTACTAACCTAATGACCTGAAGGATCGGAGCACCGCAGACAGCATAATTTGACGTGTTGCA
ACTTCACAGAATGATTCTATTATCTAACGGTTAATGTAGCGCGGTCCTTGGGGGTCCGCTAGATTTCGTGCATCCTACATATTTTTGGAGACCGGCATCTTAGTTAAATAA
TACGTATGTTCTCGATAGAAACAATGTTTAGAACTGCATCGTAAAGGAACTTGTAAACCAATGTTGGTAAAGGTAAATATGATCGGACCTGGTAATCCGCCGAAACTTAAC
TACGGCGTGATACTTAGAGGTCGCTAGTATATGATTATCATGATGCATTAAAGTTAGAACTGGCAATAATGCGTTTTCAAGATTATAGCTTATAACTTCTGCCATTGTTGG
GTGCACATATCTGCAAAAGTTGTCTGTGTCACTTCGTGCATGCATCGGTTAGTTATTAGAAGTATTAATAAGGAAACATTACACAGGGTTCTGAAAAATGAAGCGCACTGA
AGAAATTTCTAGTTAATGTACAATTAGTTAGAGGATTGTAATCATCACCACTAGGCGTAATTCGTGGTGCACAGCTTATAAATTAACTAAATGGGCATTTTGAGCTTTATG
TTCGGTAGCGCCTTTAACTATGAGTTAACTAATGGACATCTAGAAATGACAAGAATGTATTATAACTTCTGTTATTAAACTCCTCAAATACTCCTTGTATACATGTAAGTT
TTAACTTCATCAAACGTCTGCGGAAAGCGTCAGATAAATCTGGATACTTTTTTAATCTGAATTTCCTCTATGTATAATGGCCCAAATAGCAATCTCTATGTTAGAGAGAGC
CAAACCTTTCTAGAATTAATCCACACAGTTGTTATAGCTTGGTAGGAGAGTCCTATTCGATTAACGGGCTAAAAGTACAGTTTTATTGGAAACCCTTGCGAAAGAAGCATC
GTAGGTAAAATCATAAAAGCTTTCTGTCGTGCGAAATCCACGACCTACGAGGCTGCCACTTGCTCGCATACCCGTTTAGAATCCGAGGAGGTTTGTTATGCACAAGCTGAG
AAATCTACTAAACAAAAATGACGATTTAAGAGACGATCCAAAAAAATGTGCGAAAAAGGCATAAAAGAGCTTACCTGATTATGTTATGTTGATATTTAGATTTGGAGTTCA
ATTCCTCCGCGATTCGTAGAAACTAGAATGTTGTTCAGTTAATGTCGACCTCAGACACATGCCTAATATCATCTAAATTTAACAGATGCCTTTTGATAATCTTGAGGTAAA
GGACCCTTTGTTACTATGTTTTTTTAAGCATTGGTGAGTGTCAAACGTTCGGCCTTCCAGAGAAAAAAATCGCTGAAGTTAATGCAGAAGTCAGTCAAATGAACAGCTGTT
GATTCGTCCGCTTAATTCAATCTAGTTGCGAGATCGTGGACCAACGCTGTATCGCAGGAAAATCGCATGCTTTCCAAGAAAAATGGTTCGAAAAAGCCATTCACTTAACAT
CTCACTTTCTCATTACTTATATATATATATATATACATTATTTTTCTAAGGACAAGGGATGTAGTATCCAGGGGTCAAAATTAAACAAGAGGTACGTGCCCAAACTAGACG
CCAACGTGCTCCACCCGCTTATGGCCGTAATCAAGATGTGTCTTTTTTGCTAATCTAAAAAAGCGATTATAAACAAAGCTTTTACGCGCGGTTTCAGTGAAGAACCAATAC
TATTATTAGGCTAGCTATGGCTGCATCGAATGGGTCTGCTGGAATAATAAACGAAACTGTAGGCACTTCTCATTTAGAATGAAATATATGATTGTGTGAATTATTTGCGCG
TCGACTGCTTTGCGTTTAGCTGATGCATCCTGTAGATTTGAATTGTTCTACATGAGTGTTCCTAAATGACGTACAGAAAAGTCAATCACTGCTGAACACTGCTGCGTCGTT
CTTTGTGTGAGACAAGCAGGACTGACCTGTAAATCCAATTTCTAAAATAGAGTATAGATCCCTGAAGAAGATTCTCTTTTTTTTTTAATTGACTATTGCTCATGGTGGTAT
TTAAGGTGCGTTTTAGGGTGTGCTGTTGATTACGAGTGTCCTTGAGATTATATTTAGCAGTATAAACTATAAGGAGCAATAAACCTTCCTGTTTTGAAGACAC

B) Multiple sequence alignment:



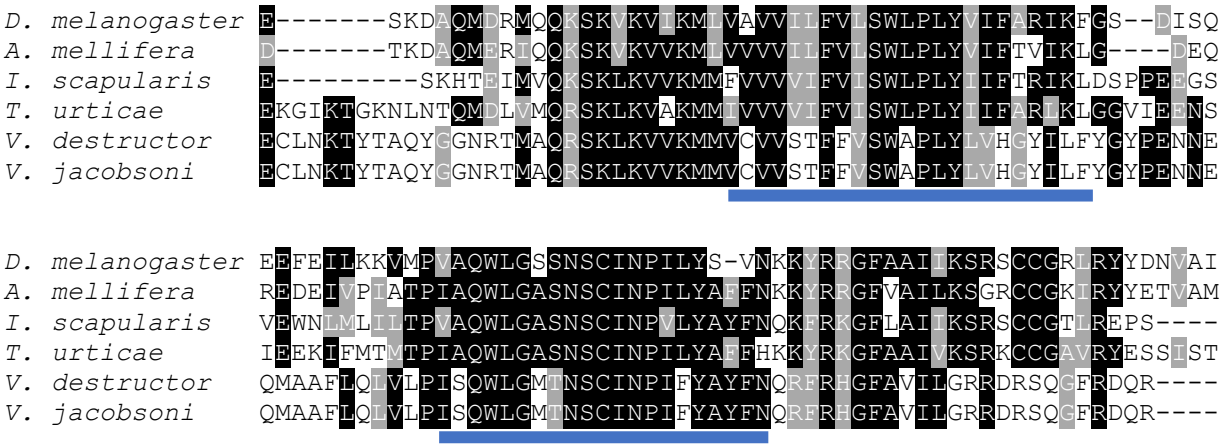


Figure 3.13 A) Gene structure for SIFamide receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of SIFamide receptor from various orthologous species across Phylum Arthropoda.

Neuropeptide F receptor

A) Gene structure:

V. destructor

```
TTGGCACGATGCTGGTGGAAATTTTAATTGAATAAAACGCAAAAAAAAACGTTGGTGTTACTATTTTGTAGTGCGGTGCGACTGAATTAGAAAAAATGCGACGACGCCGTTA
TTCTGTTGTAGCCAAAGTTCCAAACACAGATATCGATTCCGAGACTGAAGACTGAAGTAAGCACTCGCAAGAAAATAATTCAACGCCAGTGAACTTAGAAGTTCTCAAGTA
ATTGTGGCTGTGCGTAACTAAAATAGTTTTATAGGTTTGGTAGTGGCAACTAGTTTAGATGGAGCTTAATTTGTGATATGTTGACCATGCGATTGCAGAGAACGAAGAAGG

CCATCACGTGAGCACCATGTGAGACAACGCGGACTTAAACTTTGTTAACTTAATCAGGCCGCAACTGAAAATGGCGAACATGACAACCATGAAGCTCGTTGACGTCGTGAC
                                                              M  A  N  M  T  T  M  K  L  V  D  V  V  T

GTCGGTGACGCCGGCGACGGCAGCCTTCAATACAACAGAGAACCTGACTACAGAGGAAATCATGCAGCGACTTGAACAAATGATGGATATTCGAGCTATCGACGAAAATTG
  S  V  T  P  A  T  A  A  F  N  T  T  E  N  L  T  T  E  E  I  M  Q  R  L  E  Q  M  M  D  I  R  A  I  D  E  N  W

GCTTATCGGTTTATTGCTCGTCTACTCATTACTCATTATGACGGGTGCTCTTGGAAACGGACTAGTATGTTTAGCCGTCGCCCGTAAGCCAGCGATGCGGACTGCACGGAA
  L  I  G  L  L  L  V  Y  S  L  L  I  M  T  G  A  L  G  N  G  L  V  C  L  A  V  A  R  K  P  A  M  R  T  A  R  N

TATTTACATTATCAACTTGGCGATCAGCGACCTCATTTTGTGCTTATTTACTATGCCATTTTCGCTTGTTGAAATCGTCCTGAAATTTTGGCCACTCGGATTAGTTACCTG
  I  Y  I  I  N  L  A  I  S  D  L  I  L  C  L  F  T  M  P  F  S  L  V  E  I  V  L  K  F  W  P  L  G  L  V  T  C

CAAACTCGTCGCTGGCCTCGAAGCAACCTCGATTTTTGTATCAACGATCTCAATAATGGCTATTGCAATTGACCGCTACAAAGTGATCCTCTACCCGACGCGGGAGACGTT
  K  L  V  A  G  L  E  A  T  S  I  F  V  S  T  I  S  I  M  A  I  A  I  D  R  Y  K  V  I  L  Y  P  T  R  E  T  F

CAATCCACTTAGTGGCTTTTTTATGGTAGCCACCATATGGATGATTGCCCTGCTTCTTGCGACCCCACTGTTCATCTATAAAACGGTGTTTACTCATGAGGTGCCGCAGAA
  N  P  L  S  G  F  F  M  V  A  T  I  W  M  I  A  L  L  L  A  T  P  L  F  I  Y  K  T  V  F  T  H  E  V  P  Q  K

GTTTTTATTATTGCAGGTTGGCCTTCAAGCAAAAGTCGATTATTGCGTCGAGCATTGGCCGCACAATAAAGGACGGTTGTTTTTTTATTCGCTATTTACTATGGTTCTACA
  F  L  L  L  Q  V  G  L  Q  A  K  V  D  Y  C  V  E  H  W  P  H  N  K  G  R  L  F  F  Y  S  L  F  T  M  V  L  Q

GTACATGCTACCCATCGGCATTGTATCTGCGGCTTACGCTTCGATCTGTCGAAAATTACGGGACCGGATGATGGCCAGAGGGGGCTGTCAAGCTCAACTCGAAGAAAAGTT
  Y  M  L  P  I  G  I  V  S  A  A  Y  A  S  I  C  R  K  L  R  D  R  M  M  A  R  G  G  C  Q  A  Q  L  E  E  K  L

GCTTCGCGAACGGCAACGAGTTCAACGAACGAACAAACTACTAGTCGCCATTACCGTCGTCTTCATTCTAAGTTGGCTACCACTCAATATTCTCAACACTTGGTTCGACTA
  L  R  E  R  Q  R  V  Q  R  T  N  K  L  L  V  A  I  T  V  V  F  I  L  S  W  L  P  L  N  I  L  N  T  W  F  D  Y

TTTTGAGACACAGGATAGCAAACTCGACGTGCAATTTCGGGTCATGTTTGCCTTCTGTCACATGTGTGGAATGAGCTCGGCCTGTACTAACCCATTTCTATATGGCTGGCT
  F  E  T  Q  D  S  K  L  D  V  Q  F  R  V  M  F  A  F  C  H  M  C  G  M  S  S  A  C  T  N  P  F  L  Y  G  W  L

AAACGACAACTTTCGCAAGGAATTCCAAGAGATTCTCTCCGGGTGCTTCCCACAGTTCGTTGCAAGATTTACCTCTCGCCAGGATACGATCAACGAGTCTCGACTACCTTC
  N  D  N  F  R  K  E  F  Q  E  I  L  S  G  C  F  P  Q  F  V  A  R  F  T  S  R  Q  D  T  I  N  E  S  R  L  P  S
GATGCATCCCGCTGTTAGAACAATTGCTGGAAACGAAATGATACAGCTTCGTGCTTATGAAAGTCCTCCGTCATAGATATACATTAATAATCTAAATGTTGTTTAAATTAA
  M  H  P  A  V  R  T  I  A  G  N  E  M  I  Q  L  R  A  Y  E  S  P  P  S  *
AGGGGAGTGAGAAGAAAGCAGGCACAGTAGTACGTTAGATTTCCTACCAATTCCATTTCTACTATTGCTATGAAGAGGTATAGAAAAAAAAAAAAAAAAAAAAAAAGTGT
TCGTTTAACTACTAATGGCAG
```

V. jacobsoni

```
TCAAGCTAGGCTGGTCATTACGGTAGGAAAAGGTCTGAAACCAGCGGGCTTGGCACGATGCTGGTGGAAATTTTAATTGAATAAAACGCAAAAAAAAACGTTGGTGTTACTA

TTTTGTAGTGCGGTGCGACTGAATTAGAAAAAATGCGACGACGCCGTTATTCTGTTGTAGCCAAAGTTCCAAACACAGATATCGATTCCGAGACTGAAGACTGAAAGAACG
AAGAAGGCCATCACGTGAGCACCATGTGAGACAACGCGGACTTAAACTTTGTTAACTTAATCAGGCCGCAACTGAAAATGGCGAACATGACAACCATGAAGCTCGTTGACG
                                                                  M  A  N  M  T  T  M  K  L  V  D
TCGTGACGTCGGTGACGCCGGCGACGGCAGCCTTCAATACAACAGAGAACCTGACTACTGAGGAAATCATGCAGCGACTTGAACAAATGATGGATATTCGAGCTATCGATG
V  V  T  S  V  T  P  A  T  A  A  F  N  T  T  E  N  L  T  T  E  E  I  M  Q  R  L  E  Q  M  M  D  I  R  A  I  D
AAAATTGGCTTATCGGTTTATTGCTCGTCTACTCATTACTCATTATGACGGGTGCTCTTGGAAACGGACTAGTATGTTTAGCCGTCGCCCGTAAGCCAGCGATGCGGACTG
E  N  W  L  I  G  L  L  L  V  Y  S  L  L  I  M  T  G  A  L  G  N  G  L  V  C  L  A  V  A  R  K  P  A  M  R  T

CACGGAATATTTACATTATCAACTTGGCGATCAGCGACCTCATTTTGTGCTTATTTACTATGCCATTTTCGCTTGTTGAAATCGTCCTGAAATTTTGGCCGCTCGGATTAG
A  R  N  I  Y  I  I  N  L  A  I  S  D  L  I  L  C  L  F  T  M  P  F  S  L  V  E  I  V  L  K  F  W  P  L  G  L

TTACCTGCAAACTCGTCGCTGGCCTCGAAGCAACCTCGATTTTTGTATCAACGATCTCAATAATGGCTATTGCAATTGACCGCTACAAAGTGATCCTCTACCCGACGCGGG
V  T  C  K  L  V  A  G  L  E  A  T  S  I  F  V  S  T  I  S  I  M  A  I  A  I  D  R  Y  K  V  I  L  Y  P  T  R

AGACGTTCAATCCACTTAGTGGCTTTTTTATGGTAGCCACCATATGGATGATTGCCCTGCTTCTTGCGACCCCACTGTTCATCTATAAAACGGTGTTTACTCATGAGGTGC
E  T  F  N  P  L  S  G  F  F  M  V  A  T  I  W  M  I  A  L  L  L  A  T  P  L  F  I  Y  K  T  V  F  T  H  E  V

CGCAGAAGTTTTTATTATTGCAGGTTGGCCTTCAAGCAAAAGTCGATTATTGCGTCGAGCATTGGCCGCACAATAAAGGACGGTTTTTTTTATTCGCTATTTACTATGGTTC
P  Q  K  F  L  L  L  Q  V  G  L  Q  A  K  V  D  Y  C  V  E  H  W  P  H  N  K  G  R  F  F  Y  S  L  F  T  M  V

TACAGTACATGCTACCCATCGGCATTGTATCTGCGGCTTACGCTTCGATCTGTCGAAAATTACGGGACCGGATGATGGCCAGAGGGGGCTGTCAAGCTCAACTCGAAGAAA
L  Q  Y  M  L  P  I  G  I  V  S  A  A  Y  A  S  I  C  R  K  L  R  D  R  M  M  A  R  G  G  C  Q  A  Q  L  E  E

AGTTGCTTCGCGAACGGCAACGAGTTCAACGAACGAACAAACTACTAGTCGCCATTACCGTCGTCTTCATTCTAAGTTGGCTACCACTCAATATTCTCAACACTTGGTTCG
```

```
K  L  L  R  E  R  Q  R  V  Q  R  T  N  K  L  L  V  A  I  T  V  V  F  I  L  S  W  L  P  L  N  I  L  N  T  W  F

ACTATTTTGAGACACAGGATAGCAAACTCGACGTGCAATTTCGGGTCATGTTTGCCTTCTGTCACATGTGTGGAATGAGCTCGGCCTGTACTAACCCATTTCTATATGGCT
D  Y  F  E  T  Q  D  S  K  L  D  V  Q  F  R  V  M  F  A  F  C  H  M  C  G  M  S  S  A  C  T  N  P  F  L  Y  G

GGCTAAACGACAACTTTCGCAAGGAATTCCAAGAGATTCTCTCCGGGTGCTTCCCACAGTTCGTTGCAAGATTTACCTCTCGCCAGGATACGATCAACGAGTCTCGACTAC
W  L  N  D  N  F  R  K  E  F  Q  E  I  L  S  G  C  F  P  Q  F  V  A  R  F  T  S  R  Q  D  T  I  N  E  S  R  L
CTTCGATGCATCCCGCTGTTAGAACAATTGCTGGAAACGAAATGATACAGCTTCGTGGTTATGAAAGTCCTCCGTCATAGATATACATTAATAATCTAAATGTTGTTTAAA
P  S  M  H  P  A  V  R  T  I  A  G  N  E  M  I  Q  L  R  G  Y  E  S  P  P  S  *
TTAAAGAGGAGTGAGAAGAAAGCAGGCACAGTAGTACGTTAGATTTCCTACCAATTCCATTTCTACTATTGCTATGAAGAGGTATAGAAAAAAAAAAAAAAAAAAAAAAA
AAAAAGTGTTTGTTTAACTACTAATGGCAGTAGACAACAATACCGCTAATGCCAGGTAATGTTCGGCATACATCTGTACGTGTATACCATTTCAATAAATGAAATGCACAC
ATTTTCCGGCCGGGATCGTAGAGTTTACGGACTCGCGAGTCAGGTACAACAGCTGTAGGCCTTTACAATCTGGTTTTAAATCACCTTAGATTCAATAGGGCGAGAAAAGAT
TGATTTCGGTATCGGCATACCAGGTAGCAAGCAAGAAGGGAGGGCTTTTCGTGAAGATTTGTTGGTTAATTGTTTGAAAGATGCTAAAAGTAACACAATATCATATCGTTT
TTCGAAGATAGCGAACCAGATTTGAAAAGTAGTCAGGGCTAGTGAAATTTCTACGACTGCTAGCTGCGTTTATCGTACAGATCTTTCTACGAAAAGACAAAAGTAAAACCA
GCAAATAGCATGTAAAGTGAACAGCCTATACTCGCATATAAATGCTGCTTCACCTAATATCATTGTAACCATTCTCTCAAGGACATATATAATATGACCTTTTTACAAGGT
CGTGAAATTTGTTCCTTGTAAAAAATCATACGAGAACAATTTATCATCTCAACTAGATGAGTGTAATTCAGTTGGTAATTGAGACTGTGCAAGCTGGCTCTGTAGGAACAG
TTTTAGTTATTTTCAAGATACCCGCTTCTATCTGTTTTTGATTGAATCTTGTATGTGCGCACGTGTGTGTATTGCAAGGCAAAACACAACGGTCAGCGTCTTCATTTGACC
TCTTGCTAGGTATGTATGTATGTATGTATGTGCATGTGTCAAAAATTATGTAGAGGGAAGTAGCGATTCGTCAGTGTTAGCATATCCGGTGCGTTGTCGTCACTTACCTTC
GGCTCAAGTACCACCAGTTTTACATAGCACACTTGTGTGTGTAGTTTAGGTATCTTGCAGTTTCTCTATGATTCTGATAGTGCCGATGTAAATATAAGAATTCGAAATAGT
GTTTGTTAGAGAGACAATCCTGTCCTATTCGAAAGGTGCTTATCAGCTGGAAGTATACCACTATACAGGTATAGGCATAACTATTTAGTGGCGATTGAATAGTAACATTTA
TTATACCTAGGTAGATGTACAATAACTGAAACATTTTTCCCTGATGAAAATGCCACCACCAGATATTAGTTAAACGAACCTAAAAACACACATTTTAAGCAGCTAGCTTAG
GTTTCTAGAAGCCTACTTCGGCTCTCTTCCATTACACCCTGACCCTCCCGCCGCTATTTCAGGTCCATGGATTCGTTTTCCAAGCAGCAGTAATATGGACACAGGGCCATG
TGGCCGTAGGACCCCGTACAGTTAGCCCTGTGCTGGTGCGGACCGTGGCCAAATAATGAACCAATTATGACCGCAATATAGCAGAATACGTTTTGTTTTATCCGAGCACGT
TTTCATATAGCTTAATCTATGTAACTGATAGGTAAAACCGATGATTTAAATAACATTAAAAAACAACGTTTGTGTAGATTTATGACGACACATGTTTGATTGATAGTACGT
AATTGTGTATTGCAATTCTTGTGATGACAATTCAATTATTATTACTAATGTTGAATATGGCATACGAAAAAATCGAGTGATCGCATAGCTGTGATTATTTAGTACGCCCGA
ACTCACTCGCTACGGAAAAAATTCAGTCAGCGCCGTTTTACAAACAGAAATCTTATTATATACCTATAGAGATGTTTCGGACGGTCCATGTACGGTTCTGTAATACTATTA
ATTGGACAGCATTAAATCTCGTGTAAGCTCTTGAATAAAAATGTGTTCATGTTACACAGCAGTGGCGAAAACAC
```

B) Multiple sequence alignment:

```
D. melanogaster  MIISMNQTEPAQLADGEHLSGYASSSNSVRYLDDRHPLDYLDLGTVHALNITAINTSDLN
T. urticae       ------------------------------------------------------------
V. destructor    ----------------------------------------MANMTTMKLVDVVTSVTPATAAFN
V. jacobsoni     ----------------------------------------MANMTTMKLVDVVTSVTPATAAFN

D. melanogaster  ETGSRPLDPVL--IDRFLSNRAVDSPWYHMLISMYGVLIVFGALGNTLVVIAVIRKPIMR
T. urticae       ------MSEIMQKLQTYVANRAIDDTSFTYLMIAYIILIILGATGNGLVCMVVIRKPSMR
V. destructor    TTENLTTEEIMQRLEQMMDIRAIDENWLIGLLLVYSLLIMTGALGNGLVCLAVARKPAMR
V. jacobsoni     TTENLTTEEIMQRLEQMMDIRAIDENWLIGLLLVYSLLIMTGALGNGLVCLAVARKPAMR

D. melanogaster  TARNLFILNLAISDLLLCLVTMPLTLMEILSKYWPYGSCSILCKTIAMLQALCIFVSTIS
T. urticae       TPRNVFIINLAISDLLLCLFTMPFSLVEIVLKFWPLGEAT--CKLVAGLQATSIFVSTIS
V. destructor    TARNIYIINLAISDLLLCLFTMPFSLVEIVLKFWPLGLVT--CKLVAGLEATSIFVSTIS
V. jacobsoni     TARNIYIINLAISDLLLCLFTMPFSLVEIVLKFWPLGLVT--CKLVAGLEATSIFVSTIS

D. melanogaster  ITAIAFDRYQVIYPTRDSLQFVGAVTILAGIWALALLLASPLFVYKELINTDTP--ALL
T. urticae       ITAIALDRYKVIVNPNCEPCHPINTFITLSAIWITALALSAPLFIYRTVESHSIN--ILW
V. destructor    IMAIAIDRYKVILYPTRETFNPLSGFFMVATIWMIALLLATPLFIYKTVFTHEVPQKFLL
V. jacobsoni     IMAIAIDRYKVILYPTRETFNPLSGFFMVATIWMIALLLATPLFIYKTVFTHEVPQKFLL

D. melanogaster  QQIGLQDTIPYCIEDWPSRNGRFYYSIFSICVQYLVPILIVSVAYFGIYNKLKSRITVV-
T. urticae       LK-----SIDYCIEKWPIEHGRALYSIFSMIFQYVCPCIIVTVAYIRILRKLNYRMVQK-
V. destructor    LQVGLQAKVDYCVEHWPHNKGRFFYSLFTMVLQYMLPIGIVSAAYASICRKLRDRMMARG
V. jacobsoni     LQVGLQAKVDYCVEHWPHNKGRFFYSLFTMVLQYMLPIGIVSAAYASICRKLRDRMMARG

D. melanogaster  AVQASSAQRKVERGRRVKRTNCLLLISIAIIFGVSWLPLNFFNLVADIYFPFDNSQV---Y
T. urticae       KMGTTLAEKQRREKIKQRKTRLLLISISFIFGISWLPLNIVNIVADIYFPFDNSQV---Y
V. destructor    GCQAQLEEKLLRERQRVQRTNKLLVAITVVFILSWLPLNILNTWFD-YFETQDSKLDVQF
V. jacobsoni     GCQAQLEEKLLRERQRVQRTNKLLVAITVVFILSWLPLNILNTWFD-YFETQDSKLDVQF
```

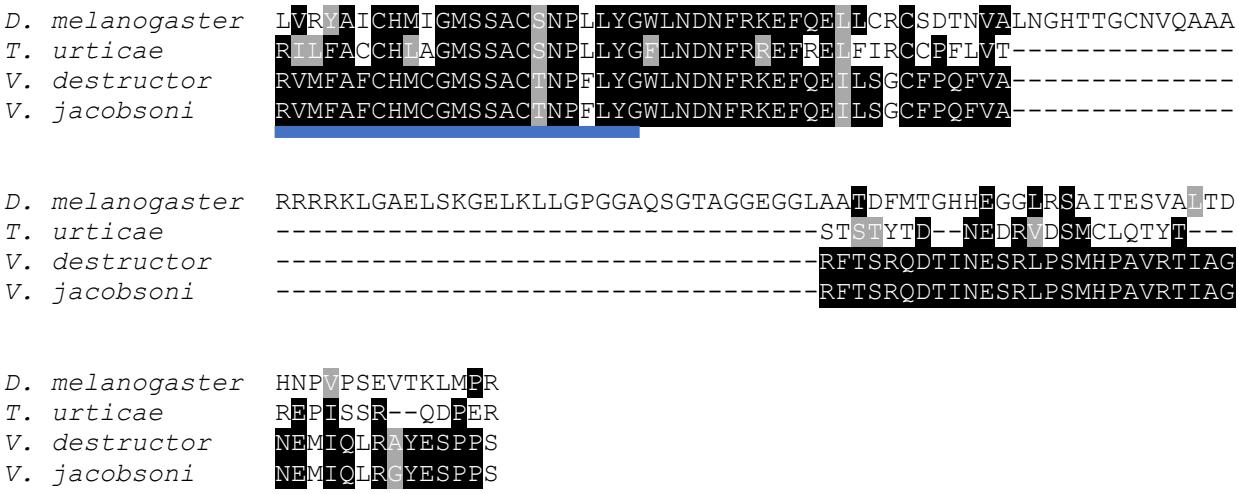


Figure 3.14 A) Gene structure for NPF receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of NPF receptor from various orthologous species across Phylum Arthropoda

Allatostatin A receptor

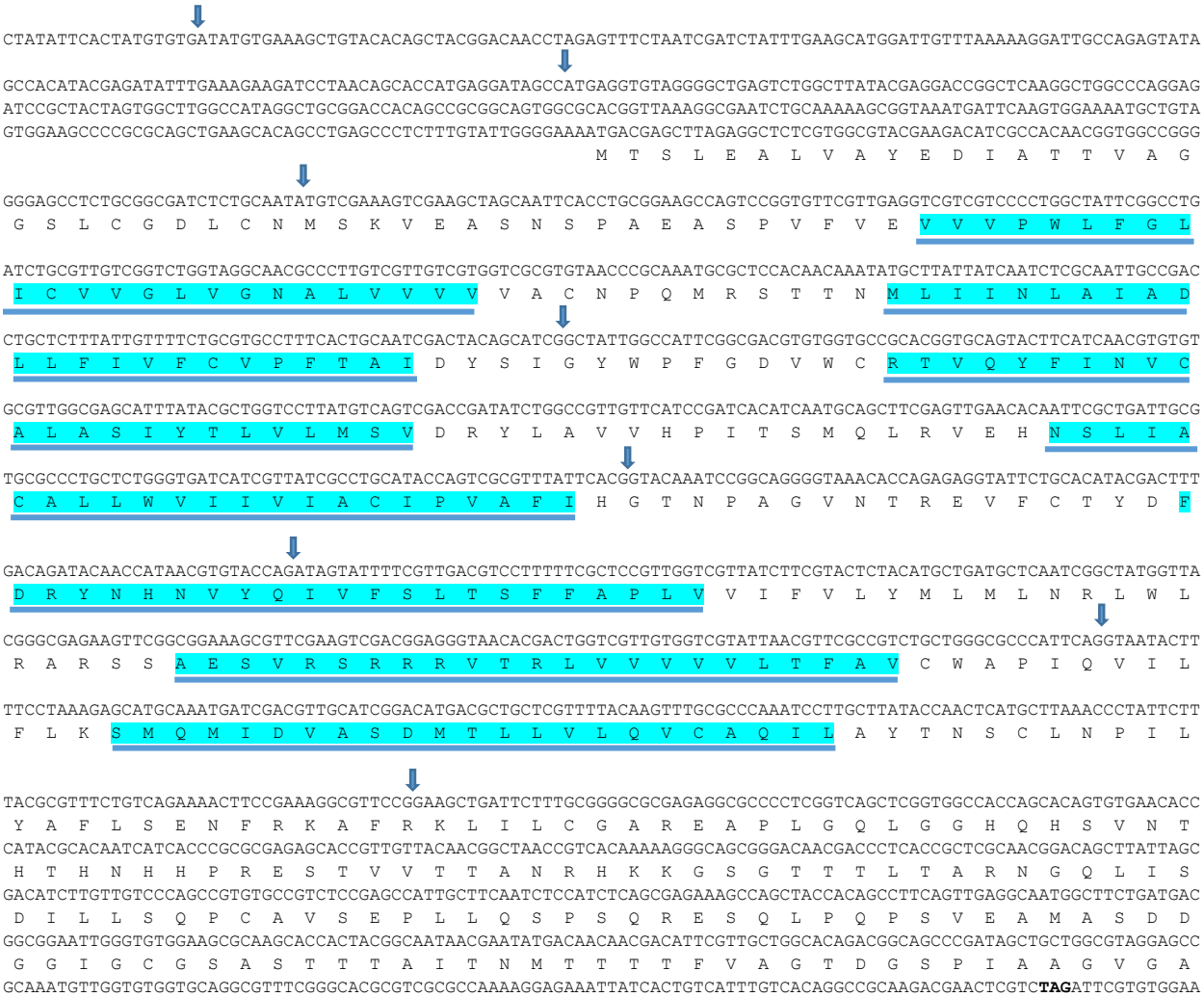
A) Gene structure

V. destructor

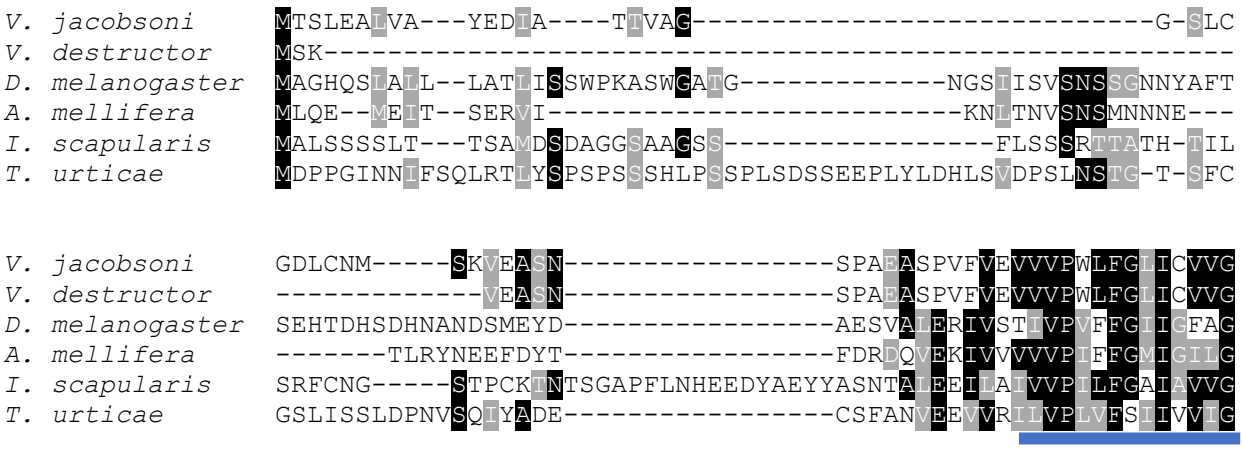
```
GACGAGCTTAGAGGCTCTCGTGGCGTACGAAGACATCGCCACAACGGTGGCCGGGGGGAGCCTCTGCGGCGATCTCTGCAATATGTCGAAAGTCGAAGCTAGCAATTCAC
                                                                                  M  S  K  V  E  A  S  N  S
CTGCGGAAGCCAGTCCGGTGTTCGTTGAGGTCGTCGTCCCCTGGCTATTCGGCCTGATCTGCGTTGTCGGTCTGGTAGGCAACGCCCTTGTCGTTGTCGTGGTCGCGTGT
P  A  E  A  S  P  V  F  V  E  V  V  V  P  W  L  F  G  L  I  C  V  V  G  L  V  G  N  A  L  V  V  V  V  V  A  C
                                                                                                             
AACCCGCAAATGCGCTCCACAACAAATATGCTTATTATCAATCTCGCAATTGCCGACCTGCTCTTTATTGTTTTCTGCGTGCCTTTCACTGCAATCGACTACAGCATCGG
 N  P  Q  M  R  S  T  T  N  M  L  I  I  N  L  A  I  A  D  L  L  F  I  V  F  C  V  P  F  T  A  I  D  Y  S  I  G
                                                                                                             
CTATTGGCCATTCGGCGACGTGTGGTGCCGCACGGTGCAGTACTTCATCAACGTGTGTGCGTTGGCGAGCATTTATACGCTGGTCCTTATGTCAGTCGACCGATATCTGG
  Y  W  P  F  G  D  V  W  C  R  T  V  Q  Y  F  I  N  V  C  A  L  A  S  I  Y  T  L  V  L  M  S  V  D  R  Y  L
                                                                                                             
CCGTTGTTCATCCGATCACATCAATGCAGCTTCGAGTTGAACACAATTCGCTGATTGCGTGCGCCCTGCTCTGGGTGATCATCGTTATCGCCTGCATACCAGTCGCGTTTA
A  V  V  H  P  I  T  S  M  Q  L  R  V  E  H  N  S  L  I  A  C  A  L  L  W  V  I  I  V  I  A  C  I  P  V  A  F
                                                                                                             
TTCACGGTACAAATCCGGCAGGGGTAAACACCAGAGAGGTATTCTGCACATACGACTTTGACAGATACAACCATAACGTGTACCAGATAGTATTTTCGTTGACGTCCTTTT
I  H  G  T  N  P  A  G  V  N  T  R  E  V  F  C  T  Y  D  F  D  R  Y  N  H  N  V  Y  Q  I  V  F  S  L  T  S  F
                                                                                                             
TCGCTCCGTTGGTCGTTATCTTCGTACTCTACATGCTGATGCTCAATCGGCTATGGTTACGGGCGAGAAGTTCGGCGGAAAGCGTTCGAAGTCGACGGAGGGTAACACGAC
F  A  P  L  V  V  I  F  V  L  Y  M  L  M  L  N  R  L  W  L  R  A  R  S  S  A  E  S  V  R  S  R  R  R  V  T  R
                                                                                                             
TGGTCGTTGTGGTCGTATTAACGTTCGCCGTCTGCTGGGCGCCCATTCAGGTAATACTTTTCCTAAAGAGCATGCAAATGATCGACGTTGCATCGGACATGACGCTGCTCG
L  V  V  V  V  V  L  T  F  A  V  C  W  A  P  I  Q  V  I  L  F  L  K  S  M  Q  M  I  D  V  A  S  D  M  T  L  L
                                                                                                             
TTTTACAAGTTTGCGCCCAAATCCTTGCTTATACCAACTCATGCTTAAACCCTATTCTTTACGCGTTTCTGTCAGAAAACTTCCGAAAGGCGTTCCGGAAGCTGATTCTTT
V  L  Q  V  C  A  Q  I  L  A  Y  T  N  S  C  L  N  P  I  L  Y  A  F  L  S  E  N  F  R  K  A  F  R  K  L  I  L
                                                                                                             
GCGGGGCGCGAGAGGCGCCCCTCGGTCAGCTCGGTGGCCACCAGCACAGTGTGAACACCCATACGCACAATCATCACCCGCGCGAGAGCACCGTTGTTACAACGGCTAACC
C  G  A  R  E  A  P  L  G  Q  L  G  G  H  Q  H  S  V  N  T  H  T  H  N  H  H  P  R  E  S  T  V  V  T  T  A  N
GTCACAAAAAGGGCAGCGGGACAACGACCCTCACCGCTCGCAACGGACAGCTTATTAGCGACATCTTGTTGTCCCAGCCGTGTGCCGTCTCCGAGCCATTGCTTCAATCTC
R  H  K  K  G  S  G  T  T  T  L  T  A  R  N  G  Q  L  I  S  D  I  L  L  S  Q  P  C  A  V  S  E  P  L  L  Q  S
CATCTCAGCGAGAAAGCCAGCTACCACAGCCTTCAGTTGAGGCAATGGCTTCTGATGACGGCGGAATTGGGTGTGGAAGCGCAAGCACCACTACGGCAATAACGAATATGA
P  S  Q  R  E  S  Q  L  P  Q  P  S  V  E  A  M  A  S  D  D  G  G  I  G  C  G  S  A  S  T  T  T  A  I  T  N  M
CAACAACGACATTCGTTGCTGGCACAGACGGCAGCCCGATAGCTGCTGGCGTAGGAGCCGCAAATGTTGGTGTGGTGCAGGCGTTTCGGGCACGCGTCGCGCCAAAAGGAG
T  T  T  T  F  V  A  G  T  D  G  S  P  I  A  A  G  V  G  A  A  N  V  G  V  V  Q  A  F  R  A  R  V  A  P  K  G
AAATTATCACTGTCATTTGTCACAGGCCGCAAGACGAACTCGTCTAG
E  I  I  T  V  I  C  H  R  P  Q  D  E  L  V  *
ATTCGTGTGGAAGAATGAAGAAAGCACTGACTAAACATAAACGAAGATTCTACGACTATCGTTAGAAATTCAACCTAATACTGTAATGATAGCCGTTTGTGCTTCCTTATC
CTCGCGATGGCTTGCTCTTTCACGTGACTGTGAATAAGATAAACACACTTCAAAAATGGAACATATCAAACAAACACAGAAAAACTACAGACGATAAATAAAAGTTATCCG
TGTTTAACGTCGCATCTTGTACGCCTCTAAAGAAAACGAAATGCATTTAAATCGCTTGAGTCAAACTGTCGTGTTTAGGAAAACCGGAAAGATTGGGATCAACACATATAT
ATACTTGTAATGCCTTTGGTGCTGTTATTATGCGAATACAACAGTGATCATTTATATAGATTAGTTATAGAAATCGTTTAATCTATTCGACATGATGTCATATGGATGATC
GCCGTAAGCAGCAGCTGTTTTGTCAGAACGACGTTGGCGACTAAAGAAACATACCACTATTGGCAACCGAAGCCCATCTAGAAACGGAAACCTTGACTGGCTCCCTAGTAG
TTTCGTCATACCAATAAAAAATCGCAAATAAGCATGCGTTTCGAGTGGTGGATCCGGTATTGCCGTTAGCATTGCTTCACATGGTTGGCCACCAGCACGAAACAAAGTAAG
TCCTTGCCCGGTCGAAGGTCGAAAAAACACTATACGTACGCCATTTTCAAAGGACACTTCGAAAGAGATTACGGCTGACTTACAGAACGTTTATGTGCATATAGAACTTTT
AGTGAACATGCATGATGTCTAGAAACGGCCCACGAAAGACACTCACCAGTACTAAGGTGCCGTCCTTGAAGCATGGCACATGGGATCATTCTATCACACTACATATAATGT
CGGATTGAATATATAGTATATTAAGTTGTACATATTCAAGTATACGCATGTAAATATTTACCAGTACGTACACAGGTATACCTTCTACGAGCTGTCAGTCAGCTTCAGAAG
GAATCAAATAGTGTACCTCGTATGAACGCATGTATGTCATCTACGGGACAAGAAGTTAAGTTCGGGAAACTTCTCTTTGTTGAAGTTCAAAGTTTCTCTCAATGCTTCTGC
TAGCGATAGACGTGAATTGCATCTCTTTCTCACAGATAGGCATATATAGCACAATGGTAAGAAGTGTCGTAATTGAAATGGGGCAGGAATGACTATGCTCTTCACCCCCTG
CCGCTCTCATTTAGACGAAATCAGAAACCCGAATTGTCCCCTATGGACCTCAACGTTGTTTTGAAAAGTTCCTAGCTGTAATAAACTTTCAACGAGATTTCTAATGTGCAA
CACGTGTATGGCAACATCATAATAACGGTAATGATTATTGCTATAACAGTATATAAATAATGTGTTTAGTTTTAATTGTCAAAATTACTATTGTAACATGTATAACACTGC
GGAACTATAACATAGCGAACAACAGGCTACAATAATTTATTGCTATCATGTACATATACAGCTGCATAGCACTGATTCCAGACCTCACGTGATGTACACTCTAAACTAATG
TATTAATGCGGTTGTTATATTATCATACCAATAATTACATTGACATAAAAACAACAATTATAAAAAACGTAATTTTTCATTTGTGATTTTATTTACTTATTACTTAATTCT
GTGTCATATATTTGTTTGACTTATATGTTTTGCGTTGATTTTGTGAACATAGGTATTGCGTTATGGAACGTAA
```

V. jacobsoni

```
AATTTTGCCGTAGCTAGCTGACTTTATTCAAAGGACTTGATGTTCCGTTGGTCGCCGTCGCTTTATCTTTGAAGCCTGCGGATCACCGACATGATGCACACATAGCATCAT
TATGAATATGGCATTCCAGTGAACACAGTCGTAGCATAGTAGCACAGGTCAAATATTTTTAGCCCATTCAGACAAAAAACTGGGCTCGGATAGATTATAAGGGAATCTGCT
TTCAATCTGTTCAACAGTATAACGAGGCACAAGGAACGCTGCTGGTTGAATAAGGGTACCAACTAAAGTGCACAGCTGGATGACGCGGCACAGTCGGCTTCGGGCAAACCA
TGAAGCGAACCTCTGCTATAGTGGAATATCTGACTAGTTAACGGATGGCGGGAATTAATATATATATATATTTCCCTTGAAATATTAGAATATTTTGTTAAAAAAGCGC
GCCAAAATATTCTGTCTTGTAGTTCATTGGGCGATATGGGCAGCCATGGTGGCTTCGAACATGCAGTTCCTCATGACCCCTGTGCACATGTTTTCTCGTCTGTGTTGTTCT
GTTTTGCTTTGTTCTATACGGATCTCATTGCCTAAAAAGCAGCCGCTCCTCGTGCTATGTTTTTTGAACAAGCCTCTGATACGTATTCTGTAATCGAATGTCAATAGGAGC
```



B) Multiple sequence alignment:



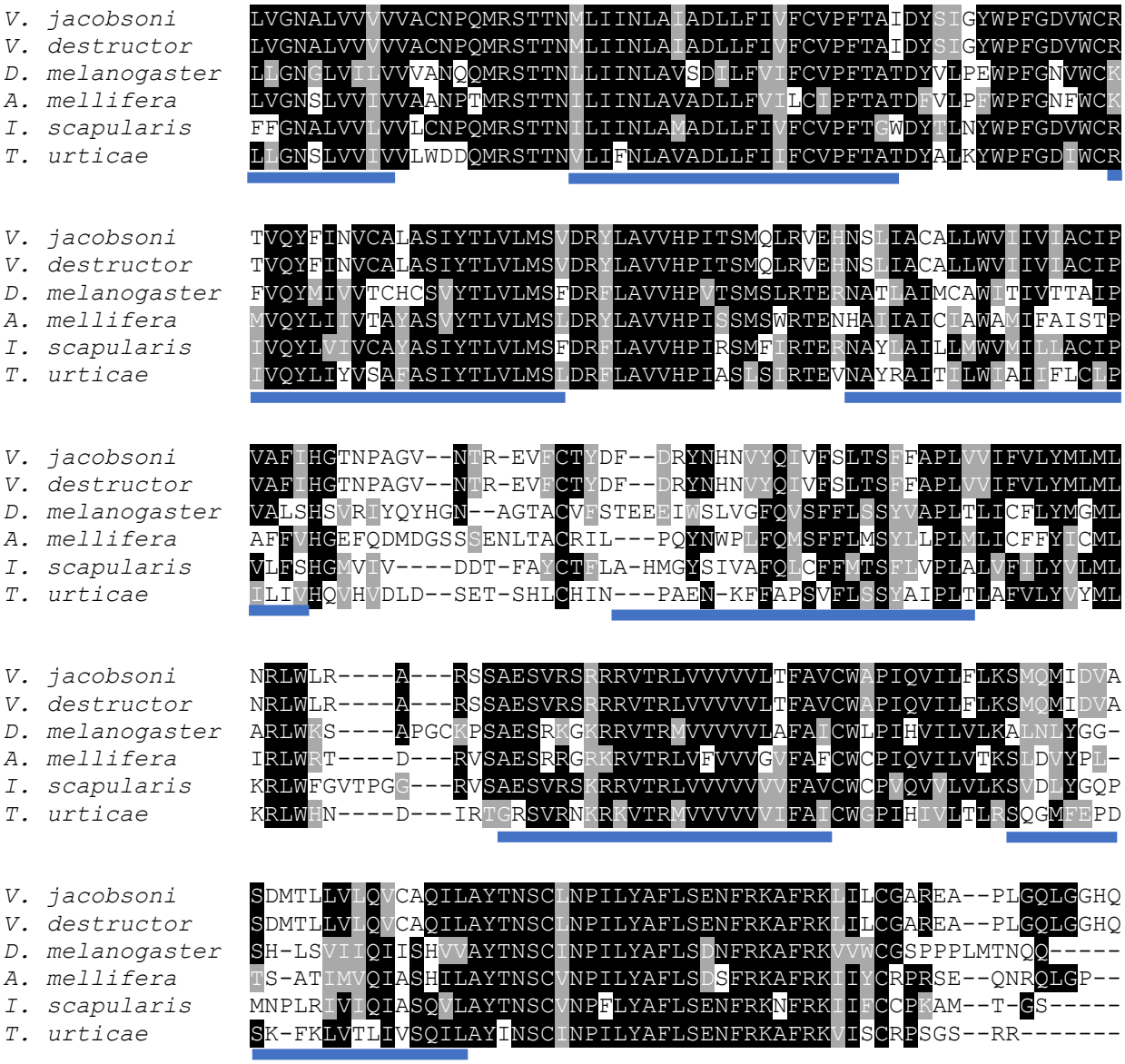
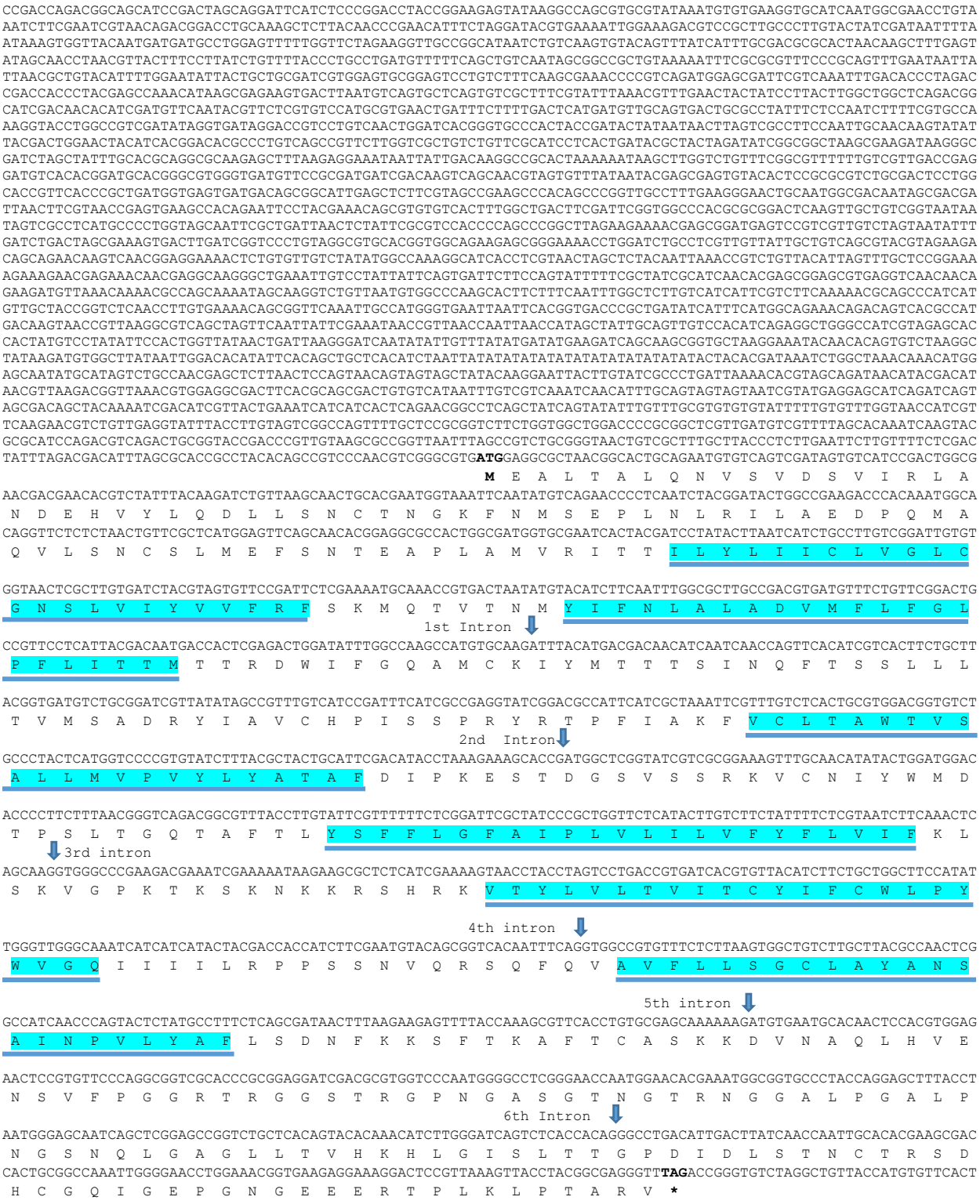


Figure 3.15 . A) Gene structure for Ast A receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of All A receptor from various orthologous species across Phylum Arthropoda.

Allatostatin C receptor

A) Gene structure

V. destructor



```
CCTCGTCATGGCTATCACAGCTAGTTATGTATCTACCTATAGAAATCATAAATTAGTATTGTTTATATCGACAATTTGTCTCACGAAGGACCGCCCGTATAAAATTACGTC
CCGCTGGTGAAGGACTTGCGTACCTAAAAAACAGTTATTTAATGGCTTTAGTGATGTTAACGGAAAAGGTTGATTATTACAGAAGGAGACAGGGTGTCTTAAAGCGAAACA
AATCTGCAATGCTAGATAGTTCTAGCATGAACTGCAGGAAAAAAGAAGTTCCGTTAGTGTCCGGACTTCTTCCGGCATCTATATACATACTAAAGATGTACAGATGTAAAG
CGATATAGTTAACGTTACTCATGGTC
```

V. jacobsoni

```
AGTGATACTATGATAGGATCAACACTGTATAGTGCCACGACTGTGAGGCAAAGGCCAAAGAAAGCTTGAAAGGGCCGACACTACGGCGACTGAGGGGTCCAAACGCCAAGA
GTCGAGGCAGTAGGCCGAGCAACCAGCCAAAACTCTAGCTGGTGTCAAGGAGAACAATTTCTTTCCAGCTCCTTTGACACGCGTAACCTGATTACCCAGTGAAATAGTGCA
AGAGAAAAAAACCAACGGAAGTATAATCTAAACACCGGGATTGTAGAGGACAGCGAGGTAACCTAATGCATTTTTCTTTATAAAGACACCAGGAAAAAGTCAAGAGAGGCT
GAAAGGAATCAACACTGCGTTGACGACTATCGACTTGGGCGTTACCCTATTCTAGGCGCCGGTTCGCGGCCGTTGGCCCTGCTGCTCTCGTTGTCGTTCGGCCTGTCTGGT
GAGAGCTCTAGTGGCGAGAGTGCGCCCTAGCAGAGAGGCCATGCGGGAAAGGGTAGAGGCCACGAGGTCCGACCAGACGGCAGCATCCGACTAGCAGGATTCATCTCCCGG
ACCTACCGGAAGAGTATAAGGCCAGCGTGCGTATAAATGTGTGAAGGTGCATCAATGGCGAACCTGTAAATCTTCGAATCGTAACAGACGGACCTGCAAAGCTCTTACAAC
CCGAACATTTCTAGGATACGTGAAAATTGGAAAGACGTCCGCTTGCCCTTGTACTATCGATAATTTTAATAAAGTGGTTACAATGATGATGCCTGGAGTTTTTGGTTCTAG
AAGGTTGCCGGCATAATCTGTCAAGTGTACAGTTTATCATTTGCGACGCGCACTAACAAGCTTTGAGTATAGCAACCTAACGTTACTTTCCTTATCTGTTTTACCCTGCCT
GATATTTTTCAGCTGTCAATAGCGGCCGCTGTAAAAATTTCGCGCGTTTCCCGCAGTTTGAATAATTATTAACGCTGTACATTTTGGAATATTACTGCTGCGATCGTGGAG
TGCGGAGTCCTGTCTTTCAAGCGAAACCCCGTCAGATGGAGCGATTCGTCAAATTTGACACCCTAGACCGACCACCCTACGAGCCAAACATAAGCGAGAAGTGACTTAATG
TCAGTGCTCAGTGTCGCTTTCGTATTTAAACGTTTGAACTACTGTCCTTACTTGGCTGGCTCAGACGGCATCGACAACACATCGATGTTCAATACGTTCTCGTGTCCATGC
GTGAACTGATTTCTTTTGACTCATGATGTTGCAGTGACTGCGCCTATTTCTCCAATCTTTTCGTGCCAAAGGTACCTGGCCGTCGATATAGGTGATAGGACCGTCCTGTCA
ACTGGATCACGGGTGCCCACCACCGATACTATAATAACTTAGTCGCCTTCCAATTGCAACAAGTATATTACGACTGGAACTACATCACGGACACGCCCTGTCAGCCGTTCT
TGGTCGCTGTCTGTTCGCATCCTCACTGATACGCTACTAGATATCGGCGGCTAAGCGAAGATAAGGGCGATCTAGCTATTTGCACGCAGGCGCAAGAGCTTTAAGAGGAAA
TAATTATTGACAAGGCCGCACTAAAAAATAAGCTTGGTCTGTTTCGGCGTTTTTTGTCGTTGACCGAGGATGTCACACGGATGCACGGGCGTGGGTGATGTTCCGCGATGA
TCGACAAGTCAGCAACGTAGTGTTTATAATACGAGCGAGTGTACACTCCGCGCGTCTGCGACTCCTGGCACCGTTCACCCGCTGATGGTGAGTGATGACAGCGGCATTGAG
CTCTTCGTAGCCGAAGCCCACAGCCCGGTTGCCTTTGAAGGGAACTGCAATGGCGACAATAGCGACGATTAACTTCGTAACCGAGTGAAGCCACAGAATTCCTACGAAACA
GCGTGTGTCACTTTGGCTGACTTCGATTCGGTGGCCCACGCGCGGACTCAAGTTGCTGTCGGTAATAATAGTCGCCTCATGCCCCTGGTAGCAATTCGCTGATTAACTCTA
TTCGCGTCCACCCCAGCCCGGCTTAGAAGAAAACGAGCGGATGAGTCCGTCGTTGTCTAGTAATATTTGATCTGACTAGCGAAAGTGACTTGATCGGTCCCTGTAGGCGTG
CACGGTGGCAGAAGAGCGGGAAAACCTGGATCTGCCTCGTTGTTATTGCTGTCAGCGTACGTAGAAGACAGCAGAACAAGTCAACGGAAGAAAACTCTGTGTTGTCTATAT
GGCCAAAGGCATCACCTCGTAACTAGCTCTACAATTAAACCGTCTGTTACATTAGTTTGCTCCGGAAAAGAAAAAACGAGAAACAACGAGGCAAGGGCTGAAATTGTCCTA
TTATTCAGTGATTCTTCCAGTATTTTTCGCTATCGCATCAACACGAGCGGAGCGTGAGGTCAACAACAGAAGATGTTAGACTAAACGCCAGCAAAATAGCAAGGTCTGTTA
ATGTGGCCCAAGCACTTCTTTCAATTTGGCTCTTGTCATCATTCGTCTTCAAAAACGCAGCCCATCATGTTGCTACCGGTCTCAACCTTGTGAAAACAGCGGTTCAAATTG
CCATGGGTGAATTAATTCACGGTGACCCGCTGATATCATTTCATGGCAGAAACAGACAGTCACGCCATGACAAGTAACCGTTAAGGCGTCAGCTAGTTCAATTATTCGAAA
TAACCGTTAACCAATTAACCATAGCTATTGCAGTTGTCCACATCAGGGGCTGGGCCATCGTAGAGCACCACTATGTCCTATATTCCACTGGTTATAACTGATTAAGGGATC
AATATATTGTTTATATGATATGAAGATCAGCAAGCGGTGCTAAGGAAATACAACACAGTGTCTAAGGCTATAAGATGTGGCTTATAATTGGACACATATTCACAGCTGCTC
ACATCTAATTATATATATATATATATATATACTACACGATAAATCTGGCTAAACAAACATGGAGCAATATGCATAGTCTGCCAACGAGCTCTTAACTCCAGTAGC
AGTAGTAGCTATACAAGGAATTACTTGTATCGCCCTGATTAAAACACGCAGCAGATAACATACGACATAACGTTAAGACGGTTAAACGTGGAGGCGACTTCACGCAGCGAC
TGTGTCATAATTTGTCGTCAAATCAACATTTGCAGTAGTAGTAATCGTATGAGGAGCATCAGATCAGTAGCGACAGCTACAAAATCGACATCGTTACTGAAATCATCATCA
CTCAGAACGGCCTCAGCTATCAGTATATTTGTTTGCGTGTGTGTATTTTTGTGTTTGGTAACCATCGTTCAAGAACGTCTATTGAGGTATTTACCTTGTAGTCGGCCAGTT
TTGCTCCGCGGTCTTCTGGTGGCTGGACCCCGCGGCTCGTTGATGTCGTTTTAGCACAAATCAAGTACGCGCATCCAGACGTCAGACTGCGGTACCGACCCGTTGTAAGCG
CCGGTTAATTTAGCCGTCTGCGGGTAACTGTCGCTTTGCTTACCCTCTTGAATTCTTGTTTTCTCGACTATTTAGACGACATTTAGCGCACCGCCTACACAGCCGTCCCAA
CGTCGGGCGTG ATG GAGGCGCTAACGGCACTGCAGAATGTGTCAGTCGATAGTGTCATCCGACTGGCGAACGACGAACACGTCTATTTACAAGATCTGTTAAGCAACTGCA
            M  E  A  L  T  A  L  Q  N  V  S  V  D  S  V  I  R  L  A  N  D  E  H  V  Y  L  Q  D  L  L  S  N  C
CGAATGGTAAATTCAATATGTCAGAACCCCTCAATCTACGGATACTGGCCGAAGACCCACAAATGGCACAGGTTCTCTCTAACTGTTCGCTCATGGAGTTCAGCAACACGG
T  N  G  K  F  N  M  S  E  P  L  N  L  R  I  L  A  E  D  P  Q  M  A  Q  V  L  S  N  C  S  L  M  E  F  S  N  T
AGGCGCCACTGGCGATGGTGCGAATCACTACGATCCTATACTTAATCATCTGCCTTGTCGGATTGTGTGGTAACTCGCTTGTGATCTATGTAGTGTTCCGATTCTCGAAAA
E  A  P  L  A  M  V  R  I  T  T  I  L  Y  L  I  I  C  L  V  G  L  C  G  N  S  L  V  I  Y  V  V  F  R  F  S  K
TGCAAACCGTGACTAATATGTACATCTTCAATTTGGCGCTTGCCGACGTGATGTTTCTGTTCGGACTGCCGTTCCTCATTACGACAATGACCACTCGAGACTGGATATTTG
M  Q  T  V  T  N  M  Y  I  F  N  L  A  L  A  D  V  M  F  L  F  G  L  P  F  L  I  T  T  M  T  T  R  D  W  I  F
                  ⇓
GCCAAGCCATGTGCAAGATTTACATGACGACAACATCAATCAACCAGTTCACATCGTCACTTCTGCTTACGGTGATGTCTGCGGATCGTTATATAGCCGTTTGTCATCCGA
G  Q  A  M  C  K  I  Y  M  T  T  T  S  I  N  Q  F  T  S  S  L  L  L  T  V  M  S  A  D  R  Y  I  A  V  C  H  P
TTTCATCGCCGAGGTATCGGACGCCATTCATCGCTAAATTCGTTTGTCTCACTGCGTGGACGGTGTCTGCCCTACTCATGGTCCCCGTGTATCTTTACGCTACTGCATTCG
I  S  S  P  R  Y  R  T  P  F  I  A  K  F  V  C  L  T  A  W  T  V  S  A  L  L  M  V  P  V  Y  L  Y  A  T  A  F
                     ⇓
ACATACCTAAAGAAAGCACCGATGGCTCGGTATCGTCGCGGAAAGTTTGCAACATATACTGGATGGACACCCCTTCTTTAACGGGTCAGACGGCGTTTACCTTGTATTCGT
D  I  P  K  E  S  T  D  G  S  V  S  S  R  K  V  C  N  I  Y  W  M  D  T  P  S  L  T  G  Q  T  A  F  T  L  Y  S
                                                  ⇓
TTTTTCTCGGATTCGCTATCCCGCTGGTTCTCATACTTGTCTTCTATTTTCTCGTAATCTTCAAACTCAGCAAGGTGGGCCCGAAGACGAAATCGAAAAATAAGAAGCGCT
F  F  L  G  F  A  I  P  L  V  L  I  L  V  F  Y  F  L  V  I  F  K  L  S  K  V  G  P  K  T  K  S  K  N  K  K  R
CTCATCGAAAAGTAACCTACCTAGTCCTGACCGTGATCACGTGTTACATCTTCTGCTGGCTTCCATATTGGGTTGGGCAAATCATCATCATACTACGACCACCATCTTCGA
S  H  R  K  V  T  Y  L  V  L  T  V  I  T  C  Y  I  F  C  W  L  P  Y  W  V  G  Q  I  I  I  I  L  R  P  P  S  S
                        ⇓
ATGTACAGCGGTCACAATTTCAGGTGGCCGTGTTTCTCTTAAGTGGCTGTCTTGCTTACGCCAACTCGGCCATCAACCCAGTACTCTATGCCTTTCTCAGCGATAACTTTA
N  V  Q  R  S  Q  F  Q  V  A  V  F  L  L  S  G  C  L  A  Y  A  N  S  A  I  N  P  V  L  Y  A  F  L  S  D  N  F
                              ⇓
AGAAGAGTTTTACCAAAGCGTTCACCTGTGCGAGCAAAAAAGATGTGAATGCACAACTCCACGTGGAGAACTCCGTGTTCCCGGGCGGTCGCACCCGCGGAGGATCGACGC
K  K  S  F  T  K  A  F  T  C  A  S  K  K  D  V  N  A  Q  L  H  V  E  N  S  V  F  P  G  G  R  T  R  G  G  S  T

GTGGTCCCAATGGGGCCTCGGGAACCAATGGAACACGAAATGGCGGTGCCCTACCAGGAGCTTTACCTAATGGGAGCAATCAGCTCGGAGCCGGTCTGCTCACAGTACACA
R  G  P  N  G  A  S  G  T  N  G  T  R  N  G  G  A  L  P  G  A  L  P  N  G  S  N  Q  L  G  A  G  L  L  T  V  H
                        ⇓
AACATCTTGGGATCAGTCTCACCACAGGGCCTGACATTGACTTATCAACCAATTGCACACGAAGCGACCACTGCGGCCAAGTTGGGGAACCTGGAAACGGTGAAGAGGA
```

```
K  H  L  G  I  S  L  T  T  G  P  D  I  D  L  S  T  N  C  T  R  S  D  H  C  G  Q  V  G  E  P  G  N  G  E  E  E
AAGGACTCCGTTAAAGTTACCTACGGCGAGGGTTTAGACCGGGTGTCTAGGCTG
  R  T  P  L  K  L  P  T  A  R  V  *

TTACCATGTGTTCACTGCGCCAGGACCTCGTCATGGCTATCACAGCTAGTTATGTATCTACCTATAGAAATCATAAATTAGTATTGTTTATATCGACAATTTGTCTCACGA
AGGACCGCCCGTATAAAATTACGTCCCGCTGGTGAAGGACTTGCGTACCTAAAAAACAGTTATTTAATGGCTTTAGTGATGTTAACGGAAAAGGTTGATTATTACAGAAGG
AGACAGGGTGTCTTAAAGCGAAACAAATCTGCAATGCTAGATAGTTCTAGCATGAACTGCAGGAAAAAAGAAGTTCCGTTAGTGTCCGGACTTCTTCCGGCATCTATATAC
ATACTAAAGATGTACAGATGTAAAGCGATATAGTTAACGTTACTCATGGTCATGGTTTGTCCACGAGACCCAGAACAGAACGAGGGCCGCGACGCGTTCGTTGAAAAAGGT
GAATGCTGAAAGATCATTCTCGTAAAGACGATAAGCACACGGCAAAAGATGCATCTTTGAGATGAAGTTAGAAAGCACACAAAAGAGGATAGAATATACTAGTTTTCTTGT
TAACGAAAATGGTTTGACAACGTAGCGGTCGTCATAGTTTGCTATAGTCAACGGGTGGGCCGTTATTGGCACCGTCTATTACCTGTGACCTAAGACACAATGAAAGTGATT
TTTTTGCTATAACTTATCAAGAGATAATTAAATTGACGTTAGATCAACTGGGCACATTTGTGTTTAGGGACGTTAGGTTAGGTTAGGTTCACATGTCTACCTTTACACATA
GTTAATTGGCAATTATAATATAATTAGGTATTTTACTTAATGAAGCGTATAGAACCGATTAGAAACGGCGTTA
```

B) Multiple sequence alignment:

```
V. jacobsoni     MEALTALQNVSVDSVIRLANDEHV------------------------------YLQD---
V. destructor    MEALTALQNVSVDSVIRLANDEHV------------------------------YLQD---
D. melanogaster MFTWLMMD---VLQF---VKGE-------------------------------------------
A. mellifera     MNTTITST---ITSV-------------------------------------------------
I. scapularis    VPL---------ECGQREAHRAHFPALFCPKPPRVRGATENGAAAHRCAGTPVRRRRLEP
T. castaneum     MYTGGD--------V-------------------------------------------------

V. jacobsoni     ---------------------------------------------LLSNCTNGK---------
V. destructor    ---------------------------------------------LLSNCTNGK---------
D. melanogaster ------------------MTAD----------------------SEANATNW-----------
A. mellifera     ---------------------------------------------------------------
I. scapularis    RTRSPGGGSESGPPVREPSADHDQHDDAAGHRGSVGPAGSVAPSRRPASGPRTEPRLEPA
T. castaneum     ---------------------------------------------------------------

V. jacobsoni     ---FNMSEPLNL-----RILAEDP--Q-MAQVLSNCSLMEFSN---T-------EA-PL-
V. destructor    ---FNMSEPLNL-----RILAEDP--Q-MAQVLSNCSLMEFSN---T-------EA-PL-
D. melanogaster ---YNTNESLYTTELNHRWISGSSTIQ-PEESLYGTDLPTYQH---C-------IATRN-
A. mellifera     ---------INGTSIDYS---------------QNHEDSTIQN---C-------EA-DL-
I. scapularis    AGRTSAG-SHGC-----RHDAIDP--SGVARAMTNETLCQAANASGA-------AA-PLA
T. castaneum     ---------LNRRDLSRC---------------T---MNESSN---LTVDEGVGDP-GL-

V. jacobsoni     ------------------------AMVRITTILYLIICLVGLCGNSLVIYVVFRFSKMQT
V. destructor    ------------------------AMVRITTILYLIICLVGLCGNSLVIYVVFRFSKMQT
D. melanogaster ------------------------SFADLFTVVLYGFVCIIGLFGNTLVIYVVLRFSKMQT
A. mellifera     ------------------------PIISLVNQILYSIVCIVGLLGNTLVIYVVLRFSKMQT
I. scapularis    VLLSPTNCSKPDNGDQPFFTSHPEFVVEFSKLLYGLVCLVGLCGNSLVLYVVLRFSKMQT
T. castaneum     ------------------------SMTIIFQTILFAAVCLVGLVGNTLVIYVVIRFSKMQT

V. jacobsoni     VTNMYIFNLALADVMFLFGLPFLITTMTTRDWIFGQAMCKIYMTTTSINQFTSSLLLTVM
V. destructor    VTNMYIFNLALADVMFLFGLPFLITTMTTRDWIFGQAMCKIYMTTTSINQFTSSLLLTVM
D. melanogaster VTNIYILNLAVADECFLIGIPFLLYTMRICSWRFGEFMCKAYMVSTSITSFTSSIFLLIM
A. mellifera     VTNMYIVNLALADECFLIGIPFLVTTMSLRSWIFGKIMCKAYMTTTSINQFTSSIFLFIM
I. scapularis    VTNMYIFNLALADEMFLVGLPFFIATMVFKYWPFGAAMCKVYMTTTSINQFTSSLLLTVM
T. castaneum     VTNMYIVNLALADECFLIGIPFLLVTMINKQWIFGNAACKAYMISTSINQFTSSIFLCIM

V. jacobsoni     SADRYIAVCHPISSPRYRTPFIAKFVCLTAWTVSALLMVPVYLYATAFDIPKESTDGSVS
V. destructor    SADRYIAVCHPISSPRYRTPFIAKFVCLTAWTVSALLMVPVYLYATAFDIPKESTDGSVS
D. melanogaster SADRYIAVCHPISSPRYRTLHIAKVVSAIAWSTSAVLMLPVILYASTVEQED----G---
A. mellifera     SADRYIAVCHPISSPKIRTPFISKVVSLTAWATSALFMIPIFLYANAMESEK----G---
```

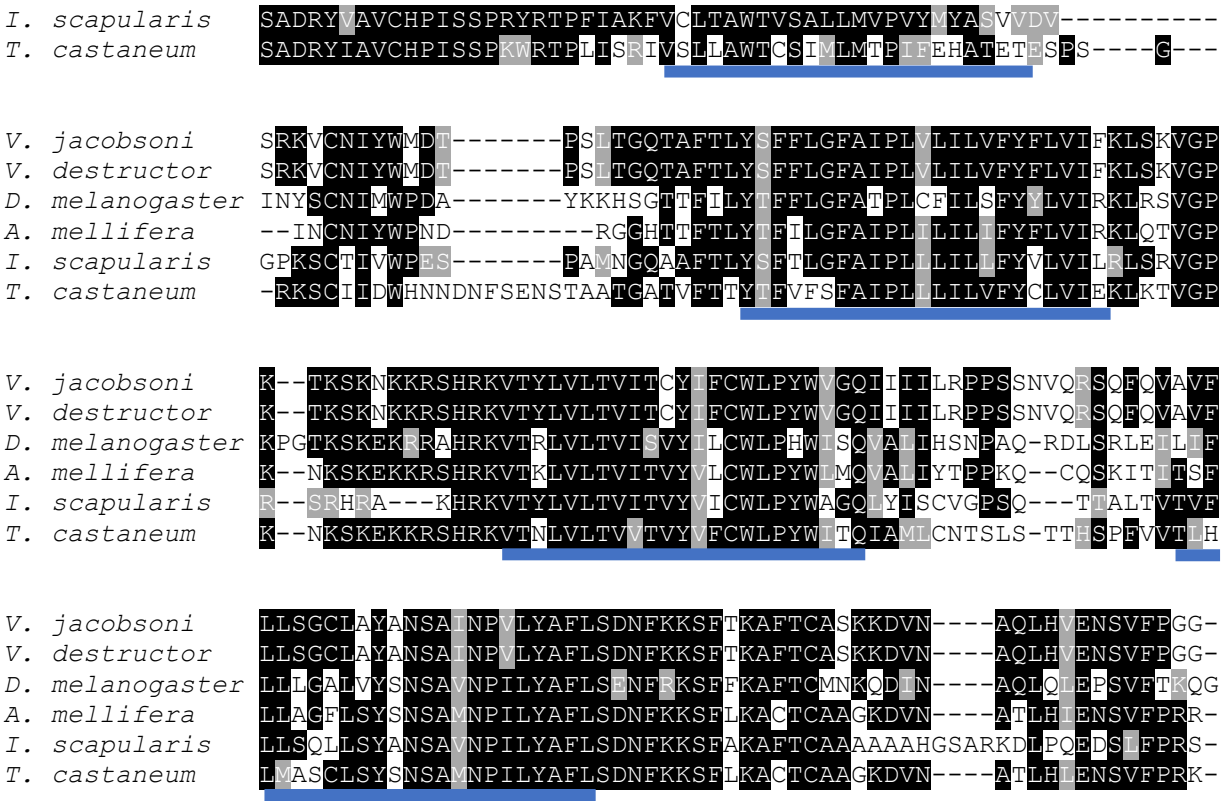
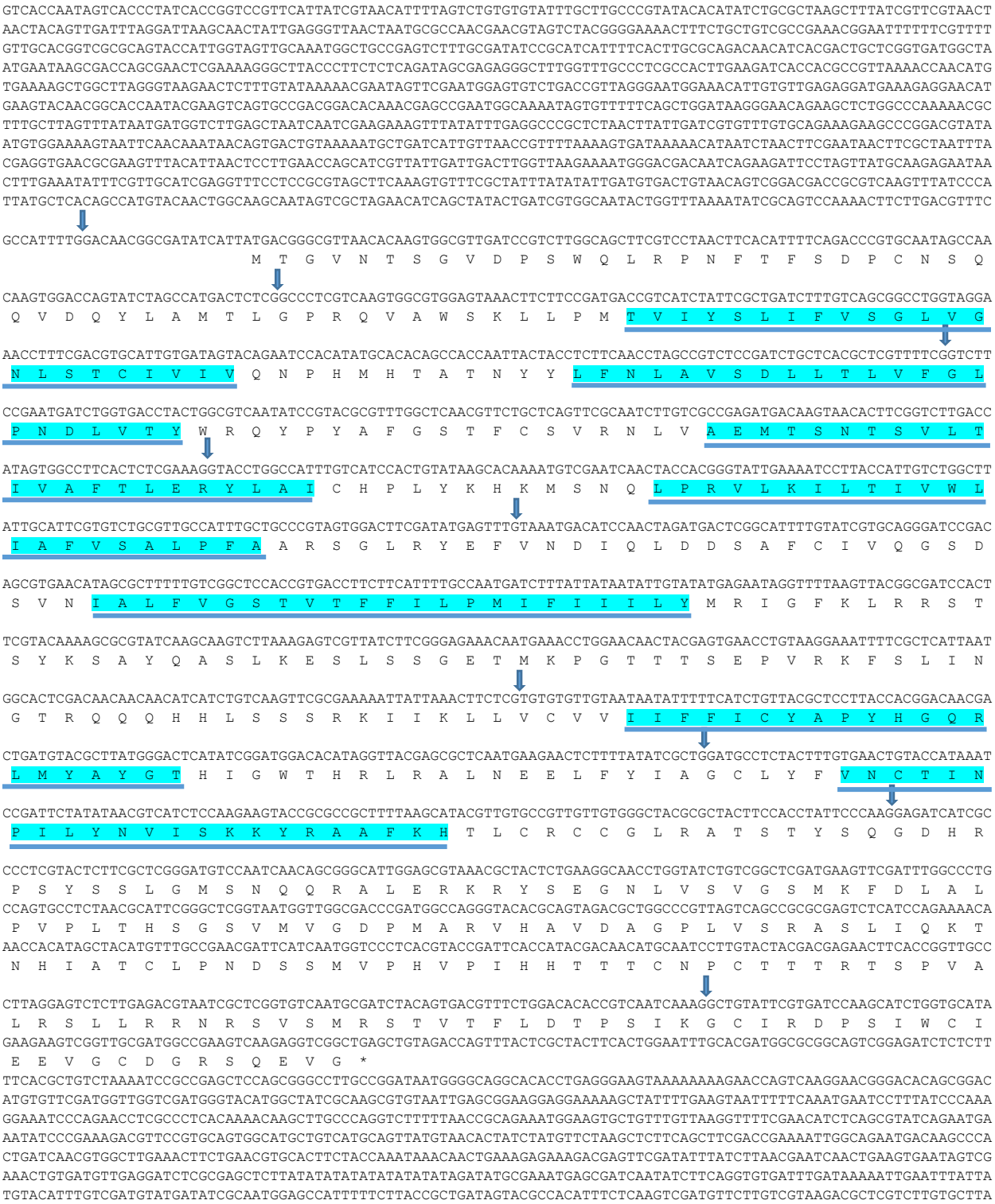


Figure 3.16 A) Gene structure for Ast C receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of All C receptor from various orthologous species across Phylum Arthropoda.

CAPA receptor

A) Gene structure:

V. destructor



```
AATCGAAACTTAGAAGCACATACACAACAAGCTTCCATGTGGCGAACAAACATTCTATCGTTCCGGAATTATCTGCTCGACTCGAAAATTGACTGGGCGATATACAGCCTA
ATACTATAAATATTTTAAATAATAAAGAAAGTTAGTGAATTTATTGCCATTCTAATTAAGTAAAATCAAATATAAACTGGGCAAAATATACAGAACCGGAAAGTAGTATAG
AGCAAGAGATATCGAACAGCAAAAAAAAAACGTGATAAATTCCTATAGAAATGGATATATACACCGAATGAATGTACATATATGTA
```

V. jacobsoni

```
CCGCAACTAGAAAGCGGCTGTTCAGAACTTGAAGGATACAGAAAATTTCATGTTGGCTTTGAAGCAGCTGAGTGCTACCAAATGACGTATTTTAACCAGGAACAACTTCGA

ATGGGACATGAGCTTCTTTAGTTAACTGAACCGTGTCTGCCGTCAGAAAGAAGCCCGGACGTATAATGTGGAAAAGTAATTCAACAAATAACAGTGACTGTAAAAATGCTG
ATCATTGTTAACCGTTTTAAAAGTGATAAAAACATAATCTAACTTCGAATAACTTCGCTAATTTACGAGGTGAACGCGAAGTTTACATTAACTCCTTGAACCAGCATCGTT
ATTGATTGACTTGGTTAAGAAAATGGGACGACAATCAGAAGATTCCTAGTTATGCAAGAGAATAACTTCGAAATATTTCGTTGCATCGAGGTTTCCTCCGCGTAGCTTCAA
AGTGTTTCGCTATTTATATATTGATGTGACTGTAACAGTCGGACGACCGCGTCAAGTTTATCCCATTATGCTCACAGCCATGTACAACTGGCAAGCAATAGTCGCTAGAAC
ATCAGCTATACTGATCGTGGCAATACTGGTTTAAAATATCGCAGTCC

AAAACTTCTTGACGTTTCGCCATTTTGGACAACGGCGATATCATTATGACGGGCGTTAACACAAGTGGCGTTGATCCGTCTTGGCAGCTTCGTCCTAACTTCACATTTTCA
                                                  M  T  G  V  N  T  S  G  V  D  P  S  W  Q  L  R  P  N  F  T  F  S

GACCCGTGCAATAGCCAACAAGTGGACCAGTATCTAGCCATGACTCTCGGCCCTCGTCAAGTGGCGTGGAGTAAACTTCTTCCGATGACCGTCATCTATTCGCTGATCTTT
 D  P  C  N  S  Q  Q  V  D  Q  Y  L  A  M  T  L  G  P  R  Q  V  A  W  S  K  L  L  P  M  T  V  I  Y  S  L  I  F

GTCAGCGGCCTGGTAGGAAACCTTTCGACGTGCATTGTGATAGTACAGAATCCACATATGCACACAGCCACCAATTACTACCTCTTCAACCTAGCCGTCTCCGATCTGCTC
 V  S  G  L  V  G  N  L  S  T  C  I  V  I  V  Q  N  P  H  M  H  T  A  T  N  Y  Y  L  F  N  L  A  V  S  D  L  L

ACGCTCGTTTTCGGTCTTCCGAATGATCTGGTGACCTACTGGCGTCAATATCCGTACGCGTTTGGCTCAACGTTCTGCTCAGTTCGCAATCTTGTCGCCGAGATGACAAGT
 T  L  V  F  G  L  P  N  D  L  V  T  Y  W  R  Q  Y  P  Y  A  F  G  S  T  F  C  S  V  R  N  L  V  A  E  M  T  S

AACACTTCGGTCTTGACCATAGTGGCCTTCACTCTCGAAAGGTACCTGGCCATTTGTCATCCACTGTATAAGCACAAAATGTCGAATCAACTACCACGGGTATTGAAAATC
 N  T  S  V  L  T  I  V  A  F  T  L  E  R  Y  L  A  I  C  H  P  L  Y  K  H  K  M  S  N  Q  L  P  R  V  L  K  I

CTTACCATTGTCTGGCTTATTGCATTCGTGTCTGCGTTGCCATTTGCTGCCCGTAGTGGACTTCGATATGAGTTTGTAAATGACATCCAACTAGATGACTCGGCATTTTGT
 L  T  I  V  W  L  I  A  F  V  S  A  L  P  F  A  A  R  S  G  L  R  Y  E  F  V  N  D  I  Q  L  D  D  S  A  F  C

ATCGTGCAGGGATCCGACAGCGTGAACATAGCGCTTTTTGTCGGCTCCACCGTGACCTTCTTCATTTTGCCAATGATCTTTATTATAATATTGTATATGAGAATAGGTTTT
 I  V  Q  G  S  D  S  V  N  I  A  L  F  V  G  S  T  V  T  F  F  I  L  P  M  I  F  I  I  I  L  Y  M  R  I  G  F

AAGTTACGGCGATCCACTTCGTACAAAAGCGCGTATCAAGCAAGTCTTAAAGAGTCGTTATCTTCGGGAGAAACAATGAAACCTGGAACAACTACGAGTGAACCTGTAAGG
 K  L  R  R  S  T  S  Y  K  S  A  Y  Q  A  S  L  K  E  S  L  S  S  G  E  T  M  K  P  G  T  T  T  S  E  P  V  R

AAATTTTCGCTCATTAATGGCACTCGACAACAACAACATCATCTGTCAAGTTCGCGAAAAATTATTAAACTTCTCGTGTGTGTTGTAATAATATTTTTCATCTGTTACGCT
 K  F  S  L  I  N  G  T  R  Q  Q  Q  H  H  L  S  S  S  R  K  I  I  K  L  L  V  C  V  V  I  I  F  F  I  C  Y  A

CCTTACCACGGACAACGACTGATGTACGCTTATGGGACTCATATCGGATGGACACATAGGTTACGAGCGCTCAATGAAGAACTCTTTTATATCGCTGGATGCCTCTACTTT
 P  Y  H  G  Q  R  L  M  Y  A  Y  G  T  H  I  G  W  T  H  R  L  R  A  L  N  E  E  L  F  Y  I  A  G  C  L  Y  F

GTGAACTGTACCATAAATCCGATTCTATATAACGTCATCTCCAAGAAGTACCGCGCCGCTTTTAAGCATACGTTGTGCCGTTGTTGTGGGCTACGCGCTACTTCCACCTAT
 V  N  C  T  I  N  P  I  L  Y  N  V  I  S  K  K  Y  R  A  A  F  K  H  T  L  C  R  C  C  G  L  R  A  T  S  T  Y

TCCCAAGGAGATCATCGCCTCTCGTACTCTTCGCTCGGGATGTCCAATCAACAGCGGGCATTGGAGCGTAAACGCTACTCTGAAGGCAACCTGGTATCTGTCGGCTCGATG
 S  Q  G  D  H  R  L  S  Y  S  S  L  G  M  S  N  Q  Q  R  A  L  E  R  K  R  Y  S  E  G  N  L  V  S  V  G  S  M
AAGTTCGATTTGGCCCTGCCAGTGCCTCTAACGCATTCGGGCTCGGTAATGGTTGGCGACCCGATGGCCAGGGTACACGCAGTAGACGCTGGCCCGTTAGTCAGCCGCGCG
 K  F  D  L  A  L  P  V  P  L  T  H  S  G  S  V  M  V  G  D  P  M  A  R  V  H  A  V  D  A  G  P  L  V  S  R  A
AGTCTCATCCAGAAAACAAACCACATAGCTACATGTTTGCCGAACGATTCATCAATGGTCCCTCACGTACCGATTCACCATACGACAACATGCAATCCTTGTACTACGACG
 S  L  I  Q  K  T  N  H  I  A  T  C  L  P  N  D  S  S  M  V  P  H  V  P  I  H  H  T  T  T  C  N  P  C  T  T  T

AGAACTTCACCGGTTGCCCTTAGGAGTCTCTTGAGACGTAATCGCTCGGTGTCAATGCGATCTACAGTGACGTTTCTGGACACACCGTCAATCAAAGGCTGTATTCGTGAT
 R  T  S  P  V  A  L  R  S  L  L  R  R  N  R  S  V  S  M  R  S  T  V  T  F  L  D  T  P  S  I  K  G  C  I  R  D
CCAAGCATCTGGTGCATAGAAGAAGTCGGTTGCGATGGCCGAAGTCAAGAGGTCGGCTGAGCTGTAGACCAGTTTACTCGCTACTTCACTGGAATTTGCACGATGGCGCGG
 P  S  I  W  C  I  E  E  V  G  C  D  G  R  S  Q  E  V  G  *
CAGTCAGAGATCTCTCTTTTCACGCTGTCTAAAATCCGCCGAGCTCCAGCGGGCCTTGCCGGATAATGGGGCAGGCACACCTGAGGGAAGTAAAAAAAAGAACCAGTCAAG
GAACGGGACACAGCGGACATGTGTTCGATGGTTGGTCGATGGGTACATGGCTATCGCAAGCGTGTAATTGAGCGGAAGGAGGAAAAAGCTATTTTGAAGTAATTTTTCAAA
TGAATCCTTTATCCCAAAGGAAATCCCAGAACCTCGCCCTCACAAAACAAGCTTGCCCAGGTCTTTTTAACCGCAGAAATGGAAGTGCTGTTTGTTAAGGTTTTCGAACAT
CTCAGCGTATCAGAATGAAATATCCCGAAAGACGTTCCGTGCAGTGGCATGCTGTCATGCAGTTATGTAACACTATCTATGTTCTAAGCTGTTCAGCTTCGACCGAAAATT
GGCAGAATGACAAGCCCACTGATCAACGTGGCTTGAAACTTCTGAACGTGCACTTCTACCAAATAAATAACTGAAAGAGAAAGACGAGTTCGATATTTATCTTAACGAATC
AACTGAAGTGAATAGTCGAAACTGTGATGTTGAGGATCTCGCGAGCTCTTATATATATATATATATAGATATGCGAAATGAGCGATCAATATCTTCAGGTGTGATTTGATA
AAAATTGAATTTATTATGTACATTTGTCGATGTATGATATCGCAATGGAGCCATTTTTCTTACCGCTGATAGTACGCCACATTTCTCAAGTCGATGTTCTTGTCGTAAGAC
```

GCTCGTCTTGTGCTTAACTCGAAACTTAGAAGCACATACACAACAAGCTTCCATGTGGCGAACAAACATTCTATCGTTCCGGAATTATCTGCTCGACTCGAAAATTGACTG
GGCGACATACAGCCTAATACTATAAATATTTTAAATAATAAAGAAAGTTAGTGAATTTATTGCCATTCTAATTAAGTAAAATCAAATATAAACTGGGCAAAATATACAGAA
CCGGAAAGGAGTATAGAGCAAGAGATATCGAACAGCAAAAAAAAACGTGATAAATTCCTATAGAAATGGATATATACACCGAATGAATGTACATATGTATGGCAAGACGAC
AAATAAAAACAATATAAACAAAGGTA

B) Multiple Sequence Alignment

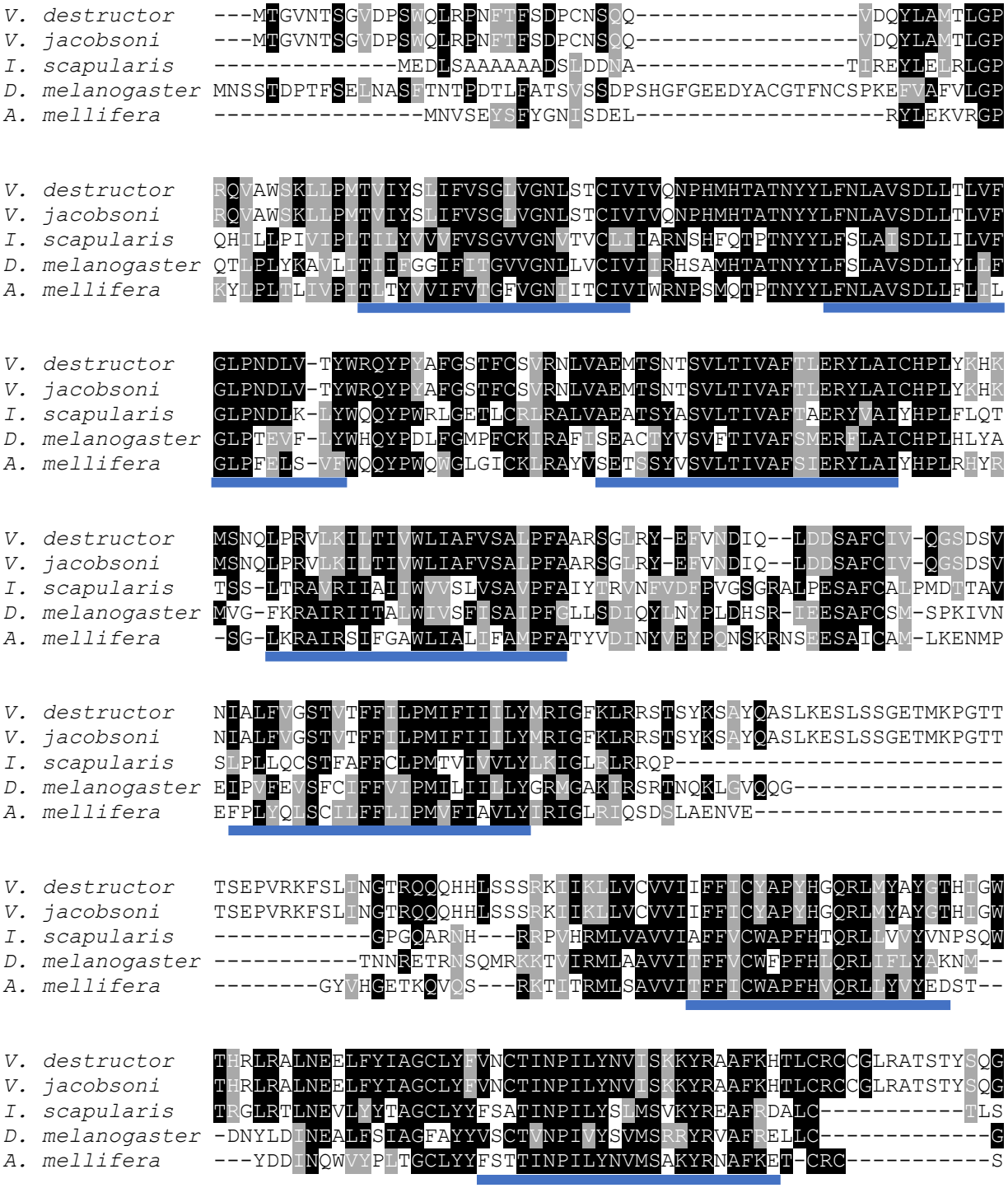


Figure 3.17 A) Gene structure for CAPA receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of CAPA receptor from various orthologous species across Phylum Arthropoda.

Myosuppressin receptor

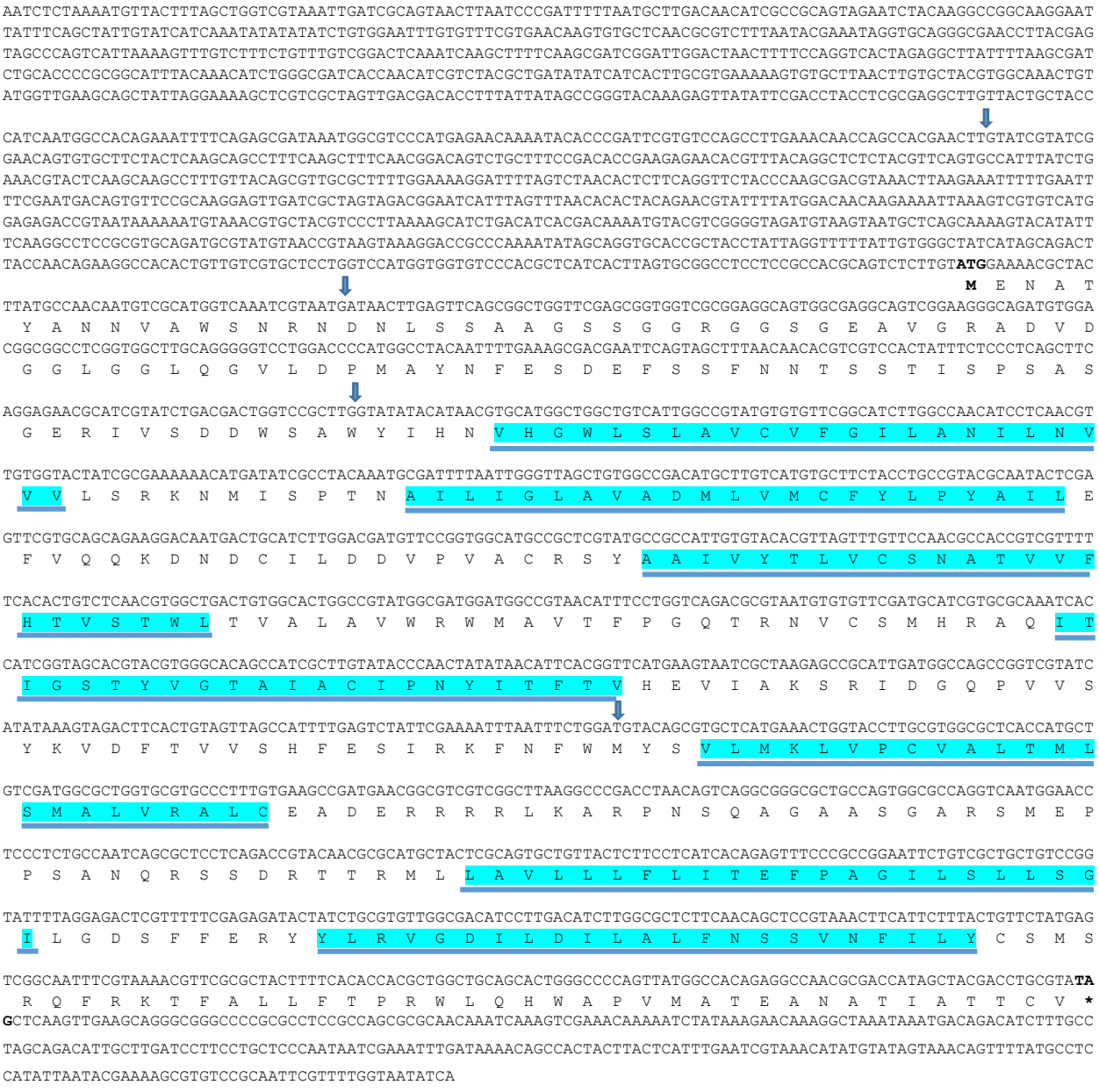
A) Gene structure

V. destructor

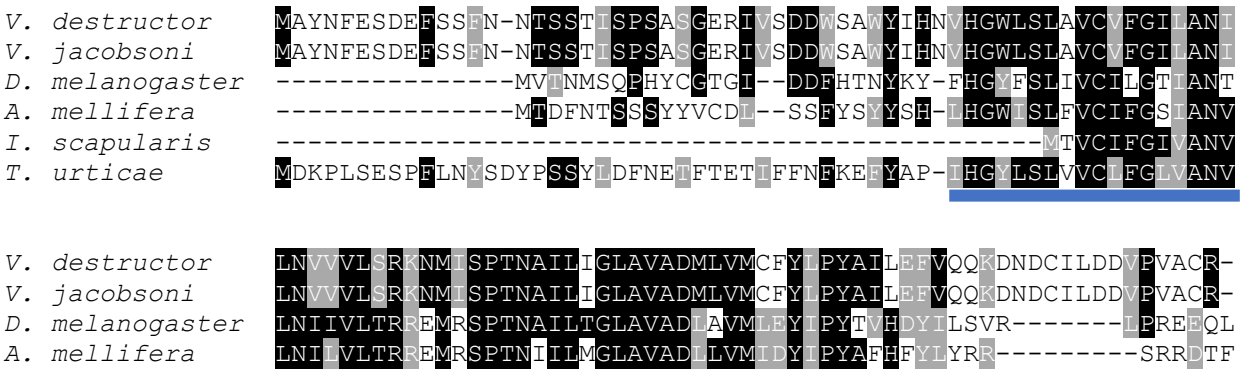
```
AAATAACCTGTTGGTTATAAGAAGACAAAAAAAGCCGAGCTAATATCTAAGATGCCTTCCGAAAAGTAACATCTATATTGTCTCATTATTTTCTGGTTCAACTGACCCTAT
CACATATATCGCGTCAAGTAACGAGAGAAGTAGCGGACCATAAAGTACTGAAAATTGAAGGTTCAAAAACATCGCTCAATGTGGTTACGTAGCCACCAGTCAGGAGCAGAA
ATATAAAATAACATAGGGAAATACCATATAGTACGAAAATAACTTTTTTCAAGTCAATTACCCAATGTTTTTTGAAAAACCGAGATGATCTAATTAGTCATTGCGAGTGAA
AATTCTACGTAGCTGATACATACTTTTCGCATCCTAACACATTAAGTGGACTTGCTTATGCAGTAATTTCCGTGGATTGCCTTTTACAGTTCTATTCAAGATCACTAAAAA
AACATCTCGTGAAACATAGTGGTGAAAAACAGATGAAACTGGTCTGTAGCTGCTTTTGCGCGTCTGCTGTGTCGTCGGCTGCTGCGGGTGATTGGATACGTTTGTGTGTGT

ACGGGTGCGCCGGTTCGTGAATATCAGATATTATTTATCAAATAAGAATTAAGGTGTATGTCTTAATGAAAAATCAATCAATGAGTGGCCAGCGTATGAATATGTAAGTAA

ATTCCAGATTTGTGTGTTCAGACAATTTTGGACTTGAGATCATGACGAAGTTCACAGTATCGTATCGGAACAGTGTGCTTCTACTCAAGCAGCCTTTCAAGCTTTCAACGG
ACAGTCTGCTTTCCGACACCGAAGAGAACACGTTTACAGGCTCTCTACGTTCAGTGCCATTTATCTGAAACGTACTCAAGCAAGCCTTTGTTACAGCGTTGCGCTTTTGGA
AAAGGATTTTAGTCTAACACTCTTCAGGTTCTACCCAAGCGACGTAAACTTAAGAAATTTTTGAATTTTCGAATGACAGTGTTCCGCAAGGAGTTGATCGCTAGTAGACGG
AATCATTTAGTTTAACACACTACAGAACGTATTTTATGGACAACAAGAAAATTAAAGTCGTGTCATGGAGAGACCGTAATAAAAAATGTAAACGTGCTACGTCCCTTAAAA
GCATCTGACATCACGACAAAATGTACGTCGGGGTAGATGTAAGTAATGCTCAGCAAAAGTACATATTTCAAGGCCTCCGCGTGCAGATGCGTATGTAACCGTAAGTAAAGG
ACCGCCCAAAATATAGCAGGTGCGCCGCTACCTATTAGGTTTTTATTGTGGGCTATCATAGCAGACTTACCAACAGAAGGCCACACTGTTGTCGTGCTCCTGGTCCATGGT
GGTGTCCCACGCTCATCACTTAGTGCGGCCTCCTCCGCCACGCAGTCTCTTGT ATG GAAAACGCTACTTATGCCAACAATGTCGCATGGTCAAATCGTAATGATAACTTGA
                                                    M  E  N  A  T  Y  A  N  N  V  A  W  S  N  R  N  D  N  L
GTTCAGCGACTGGTTCGAGCGGTGGTCGCGGAGGCAGTGGCGAGGCAGTCGGAAGGGCAGATGTGGACGGCGGCCTCGGTGGCTTGCAGGGGGTCCTGGACCCCATGGCCT
S  S  A  T  G  S  S  G  G  R  G  G  S  G  E  A  V  G  R  A  D  V  D  G  G  L  G  G  L  Q  G  V  L  D  P  M  A

ACAATTTTGAAAGCGACGAATTCAGTAGCTTTAACAACACGTCGTCCACTATTTCTCCCTCAGCTTCAGGAGAACGCATCGTATCTGACGACTGGTCCGCTTGGTATATAC
Y  N  F  E  S  D  E  F  S  S  F  N  N  T  S  S  T  I  S  P  S  A  S  G  E  R  I  V  S  D  D  W  S  A  W  Y  I
ATAACGTGCATGGCTGGCTGTCATTGGCCGTATGTGTGTTCGGCATCTTGGCCAACATCCTCAACGTTGTGGTACTATCGCGAAAAAACATGATATCGCCTACAAATGCGA
H  N  V  H  G  W  L  S  L  A  V  C  V  F  G  I  L  A  N  I  L  N  V  V  V  L  S  R  K  N  M  I  S  P  T  N  A

TTTTAATTGGGTTAGCTGTGGCCGACATGCTTGTCATGTGCTTCTACCTGCCGTACGCAATACTCGAGTTCGTGCAGCAGAAGGACAATGACTGCATCTTGGACGATGTTC
I  L  I  G  L  A  V  A  D  M  L  V  M  C  F  Y  L  P  Y  A  I  L  E  F  V  Q  Q  K  D  N  D  C  I  L  D  D  V

CGGTGGCATGCCGCTCGTATGCCGCCATTGTGTACACGTTAGTTTGTTCCAACGCCACCGTCGTTTTTCACACTGTCTCAACGTGGCTGACTGTGGCACTGGCCGTATGGC
P  V  A  C  R  S  Y  A  A  I  V  Y  T  L  V  C  S  N  A  T  V  V  F  H  T  V  S  T  W  L  T  V  A  L  A  V  W

GATGGATGGCCGTAACATTTCCTGGTCAGACGCGTAATGTGTGTTCGATGCATCGTGCGCAAATCACCATCGGTAGCACGTACGTGGGCACAGCCATCGCTTGTATACCCA
R  W  M  A  V  T  F  P  G  Q  T  R  N  V  C  S  M  H  R  A  Q  I  T  I  G  S  T  Y  V  G  T  A  I  A  C  I  P

ACTATATAACATTCACGGTTCATGAAGTAATCGCTAAGAGCCGCATTGATGGCCAGCCGGTCGTATCATATAAAGTAGACTTCACTGTAGTTAGCCATTTTGAGTCTATTC
N  Y  I  T  F  T  V  H  E  V  I  A  K  S  R  I  D  G  Q  P  V  V  S  Y  K  V  D  F  T  V  V  S  H  F  E  S  I

GAAAATTTAATTTCTGGATGTACAGCGTGCTCATGAAACTGGTACCTTGCGTGGCGCTCACCATGCTGTCGATGGCGCTGGTGCGTGCCCTTTGTGAAGCCGATGAACGGC
R  K  F  N  F  W  M  Y  S  V  L  M  K  L  V  P  C  V  A  L  T  M  L  S  M  A  L  V  R  A  L  C  E  A  D  E  R

GTCGTCGGCTTAAGGCCCGACCTAACAGTCAGGCGGGCGCTGCCAGTGGCGCCAGGTCAATGGAACCTCCCTCTGCCAATCAGCGCTCCTCAGACCGTACAACGCGCATGC
R  R  R  L  K  A  R  P  N  S  Q  A  G  A  A  S  G  A  R  S  M  E  P  P  S  A  N  Q  R  S  S  D  R  T  T  R  M

TACTCGCAGTGCTGTTACTCTTCCTCATCACAGAGTTTCCCGCCGGAATTCTGTCGCTGCTGTCCGGTATTTTAGGAGACTCGTTTTTCGAGAGATACTATCTGCGTGTTG
L  L  A  V  L  L  L  F  L  I  T  E  F  P  A  G  I  L  S  L  L  S  G  I  L  G  D  S  F  F  E  R  Y  Y  L  R  V

GCGACATCCTTGACATCTTGGCGCTCTTCAACAGCTCCGTAAACTTCATTCTTTACTGTTCTATGAGTCGGCAATTTCGTAAAACGTTCGCGCTACTTTTCACACCACGCT
G  D  I  L  D  I  L  A  L  F  N  S  S  V  N  F  I  L  Y  C  S  M  S  R  Q  F  R  K  T  F  A  L  L  F  T  P  R

GGCTGCAGCACTGGGCCCCAGTTATGGCCACAGAGGCCAACGCGACCATAGCTACGACCTGCGTA TAG CTCAAGTTGAAGCAGGGCGGGCCCCGCGCCTCCGCCAGCGCGC
W  L  Q  H  W  A  P  V  M  A  T  E  A  N  A  T  I  A  T  T  C  V  *
AACAAATCAAAGTCGAAACAAAAATCTATAAAGAACAAAGGCTAAATAAATGACAGACATCTTTGCCTAGCAGACATTGCTTGATCCTTCCTGCTCCCAATAATCGAAATT
TGATAAAACAGCCACTACTTACTCATTTGAATCGTAAACATATGTATAGTAAACAGTTTTATGCCTCCATATTAATACGAAAAGCGTGTCCGCAATTCGTTTTGGTAATAT
CATAGCGGTTCGTTTTCATTTTATATGATGATATTTAGTGATATAAATACAATGAATACGTACGCTCGA
```

V. jacobsoni:

```
AGCCTGCGCTTATACAAGGTCACTAAGAAAAACCGTCTGCTTGAAGGAACCTTTCGTTAATTATACAAGGGACTGTTCACCGCAGGATCGGATCACTAGAGCTGCACAACA
ACCAGCAATCTTATGTTTTTTAAACCTTTTTTATGGATATAAAAGGTGGTTTTGAAAGCTGTTTACAAAATCGGAATCATACGCAATCAATAGAAATTTAAGAGGAACTCT
TAGGGAGGTAATCCCTGTATATGACAAATTTTTTTATACCGATCCTGATTGATCGTTACCAAACCAGAGTGAAAAAATGTACTATGCAGCGGATGAAGCAACGTTCAGAAT
ACAATAAGATAAATTATAACACCCGCGACACATCGTCCCATGGTAGCAAGACAAACCCTGTGAAAAAGATGACCAAAACAGACAAATTAATTGTTTGCAGATAGTCCCTAA
ATGAATTCGATCATATGGCTGTCTATAGTACGTGTTACTATTCTCATTAGGTGATGCTGGTTGTCATCAAAAAAAAAACATATATGATATCAAGAAAAACAGATACGTTAA
CAATACTTGTATTGAATGATCCTCTAAAAATTAAAAAAATATCGTTTATTTAATCAAGTAATGGAAGATGCTAACGCTACGCATTTCTCCACAGTTTTGAAATATCTGGAC
TACCGCCTCATAACATACCTATTATCCTGATAAGAAGTAAAGACGCTTTTAGCTATTTCCAAGGATTTGTATTTCATCGCGATCAACTAATGCGCTTTCAGTTTCGTGTAA
```



B) Multiple sequence alignment:



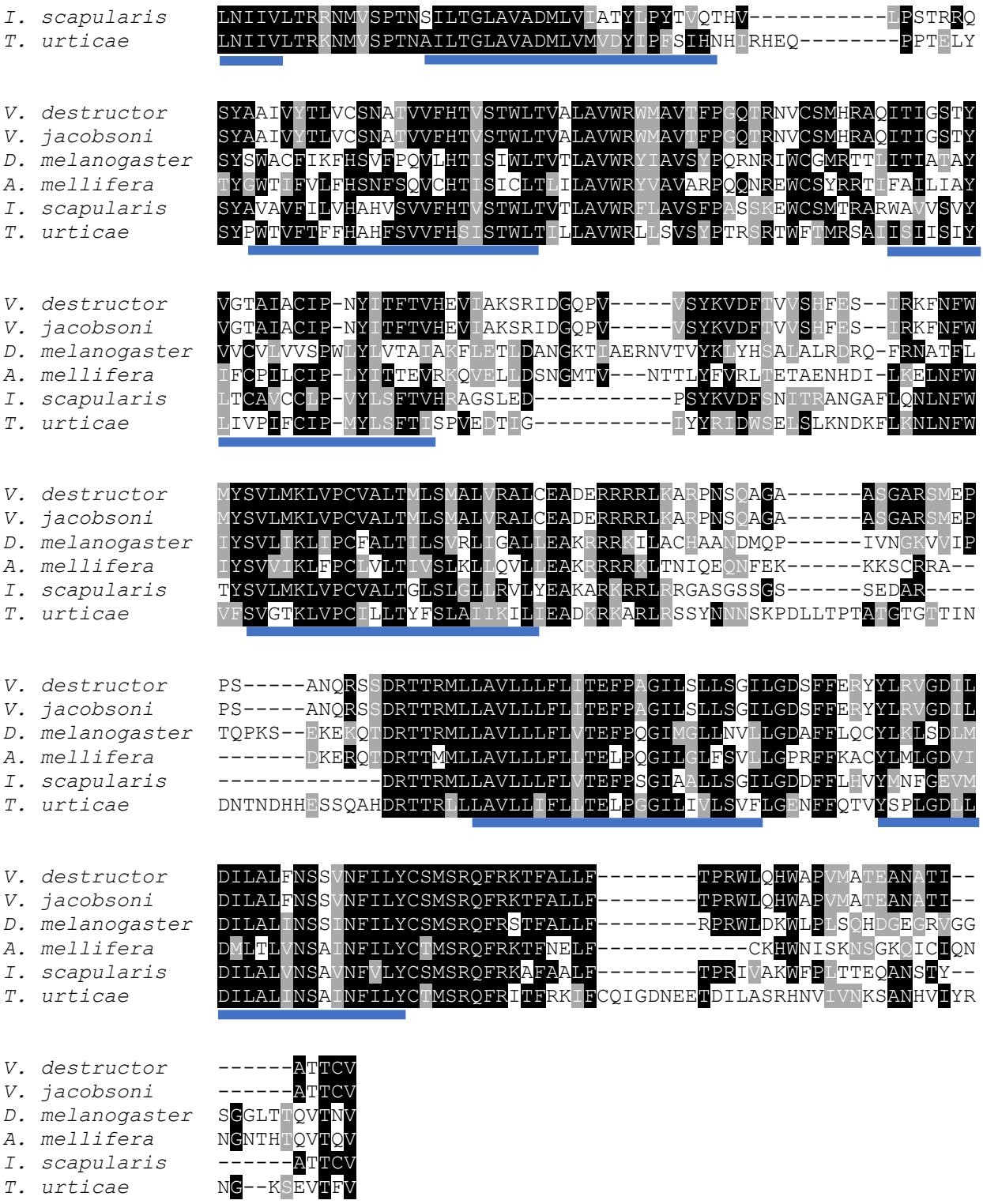
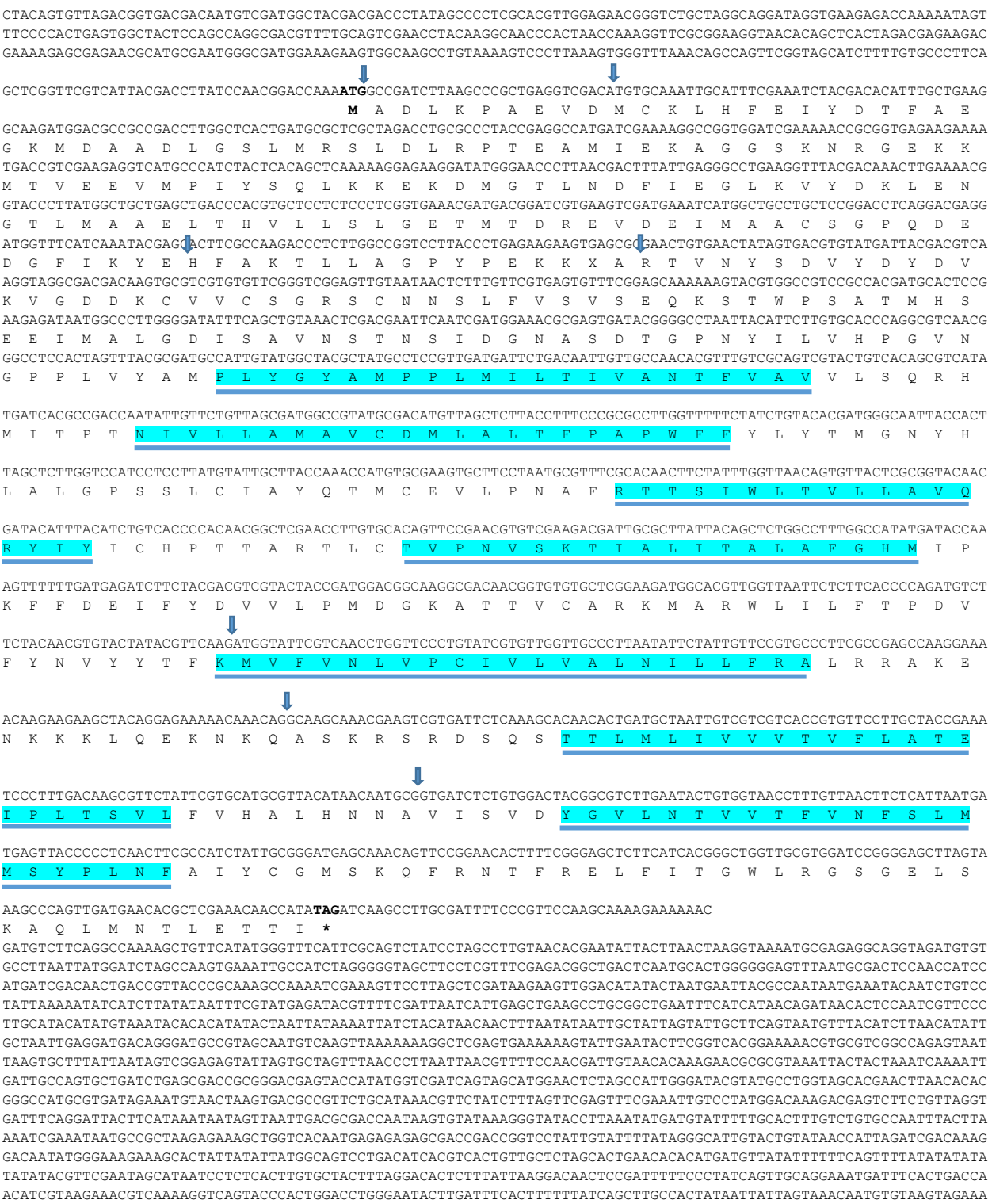


Figure 3.18 A) Gene structure for Myosuppressin receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of myosuppressin receptor from various orthologous species across Phylum Arthropoda.

Allatostatin B receptor

A) Gene structure

V. destructor



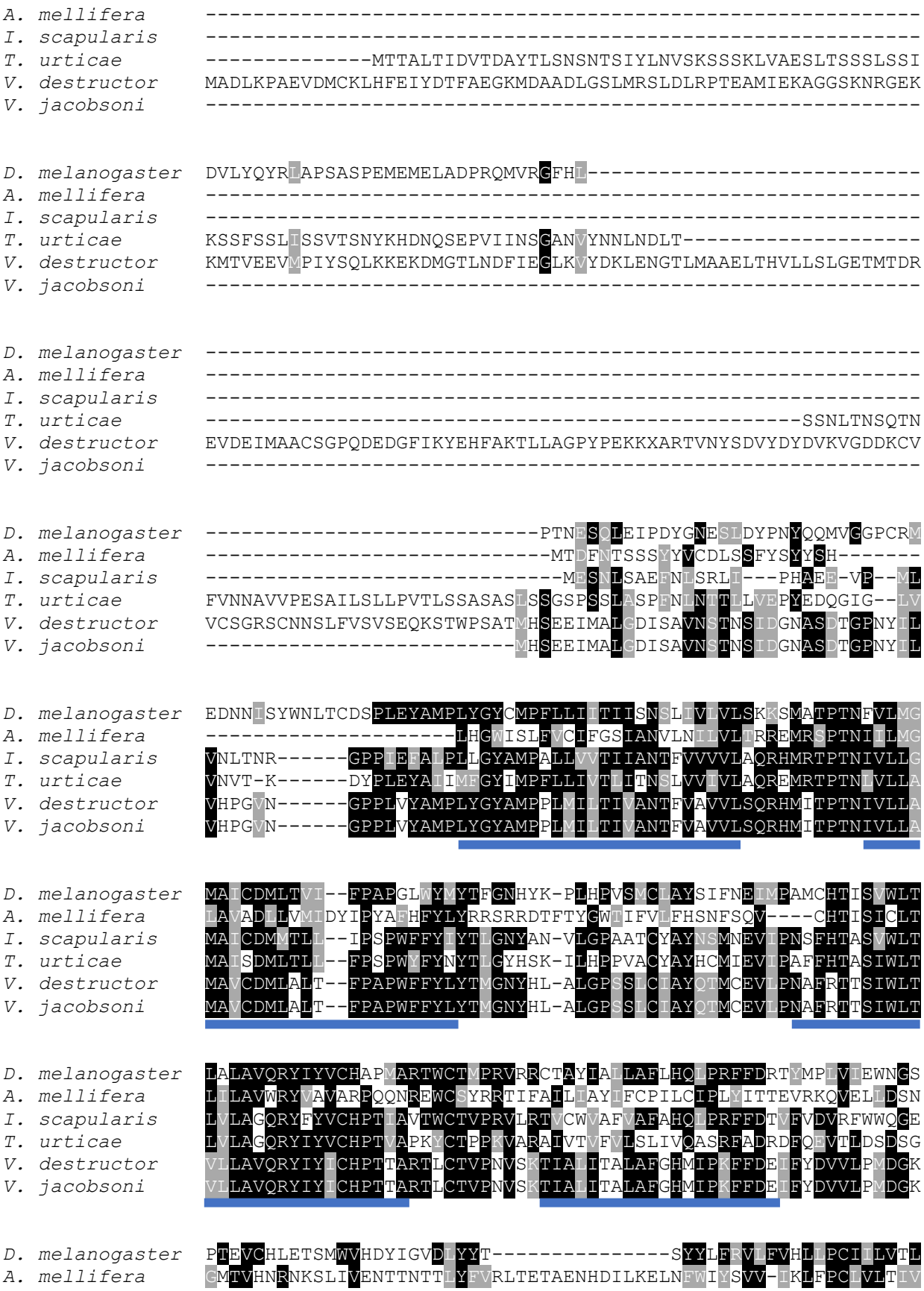
```
GAGTTACACTTTGGCTAATGAAATTCTGGTTTCTGTGATCTGTGCCAAAAATACTTATTTTTTTTTTCTATCTATTTCAAAGAACTAAAACGTTATCGCTGTTATCTGAAG
GTATGTGATGTGAAATCAGATATCACAACGGTGCAGCCTCTATACTCCCATAAGGAAAATGGGGACCGTTATTGGTTAGTAACGCGCGTAGATGCGGCATGATTAGTGTTT
TCTTCAGAAATTATGTGAGGACAATAACTGTGCTCTTCCTTGTTCTGTGGGCCTGTAAAGTAGCTCCTGGCGATTTCCTAAATTAAATAAACAAAGTACGGATTATGTTTG
TTGTTGCAAGACGATATGCTGAGTCTTAGCTGATCAACATTCATTCCGTAACTGTCAGGCTTTAGGACGTTGTTTGCAACATGCTTACGGATTTGCAAAAAAATGCTATCG
TTTACCCGTGCTTTGTTATGAAATTTTTGATGAAGCGGTGCTGTACGGTCACGGCACGATCTAAGTAGTTCGGTATGGACATAATTAGGCGATACGCTCTCAAATATTTGT
GGAAATCAATCGTCGTTTTTCTTTCCTAGCCCCCTAATCCGACCGGCGAAACAATGTAAATCGTCTAGCCGACCTGCGAAGGGTACGTCCAACGACACTTCACCCGCTGGT
ATTGGTAAATACCGCTTCCCTCTTCATGCGCTGCCCTGTTAACCGCTGTGAGCCGCAAAGGTTAGATACTGTTGGATTACGTGACAGGTAGATTTCCTGCAAGTCTTCATG
ATACGTTATCGTATTACGATGCATACAGTTGAAAGTAACAGCGTGAAAATACTGATATGAAATTAGTTCTGTGGACTAGAAGGCTCTTGTTTTGGTCATGCCATACTACAA
ATGCAAATTACATTAGACCAATATAAACTAAAATAGAAAAAAGGGAATCACCAACTATCAAAATAAACAGTAATATTTTTTAAAAACAACCCCGCTGTTTTGTCTCTAATT
TACATATGCCGATCATCGTTTCACCTTCTCGTCTTGTTTCAGACTCGATGTATCGCACGGATGCCGCTAGAAAAATGTTTTAGACTAAACAGAAGAATACAGAAACTTACT
TTTTGCATAGAAACTTGCTCTGGCTAAATATCAGCCCGCTCTTCAAACTGAAGACACACCGTCAAACCCTTATTTTCGGAGAAGTTGCTTTATTGTTTCCAATCGCTGCAT
TATACCGTAATACCAATAGCAGGCCGAAGGCTCGCTCGTTGTGACATCTAAATTATTGGCTGCTACTAATTGGGCCTTTAAAGCACAGGTAGTGCTGGTTGCGCCCGTACC
TGAAACAGATGTAAGCAAGTGATTTTAACAGATATCACTTGTGAAGAACACAAATATCATTTAGGCTTGCTTTTTTCGCGCCTTATCGCAAAGAAGAGATTTTTTCTGTTC
TATTCGAAAAAATATTTTCGATAAATAACTAAATTCTGGAGTGGATTAAGTGCCGTGTATATA
```

V. jacobsoni

```
GCTAAATGCTAGTAGTTCCGGCACCTCGAAACATAATATAGCCGCCGATATCCAGTGGGGACCCAGACCCGGGTCATCACTTGCCAACACACTTGTGGACAGCGCATTTTC
CGTTGGGGACCTGTTTTTTGGATGAGCGCAGCAAAGTGTAGGAGAGACACGGCAAGGTTCGAACCCTGAAAGGAAATGCTTGCAAATATCGAAGGAATGGCAGTAATTTAA
GCAGAACGTGGACACGGCATACTGATTTCACGCCGGTTTAGACGTTCGGTCTGATGCGGCTTACCAGTTGAATTAAATTAGCAGCGTTTCTTTGTCCTACACGTCCCGATT
AAGGCACGCGTGGATTGATGCTTGGAGGAGAGGGTCTAGTTTAGAATAGTTTTCAGAATTTCTGAACCGAGTCAATGAACTGTGAACTATAGTGACGTGTATGATTACGAC
GTCAAGGTAGGCGACGACAAGTGCGTCGTGTGTTCGGGTCGGAGTTGTAATAACTCTTTGTTCGTGAGTGTTTCGGAGCAAAAAAGTACGTGGCCGTCCGCCACGATGCAC
                                                                                                          M  H
TCCGAAGAGATAATGGCCCTTGGGGATATTTCAGCTGTAAACTCGACGAATTCAATCGATGGAAACGCGAGTGATACGGGGCCTAATTACATTCTTGTGCACCCAGGCGTC
 S  E  E  I  M  A  L  G  D  I  S  A  V  N  S  T  N  S  I  D  G  N  A  S  D  T  G  P  N  Y  I  L  V  H  P  G  V
AACGGGCCTCCACTAGTTTACGCGATGCCATTGTATGGCTACGCTATGCCTCCGTTGATGATTCTGACAATTGTTGCCAACACGTTTGTCGCAGTCGTACTGTCACAGCGT
 N  G  P  P  L  V  Y  A  M  P  L  Y  G  Y  A  M  P  P  L  M  I  L  T  I  V  A  N  T  F  V  A  V  V  L  S  Q  R

CATATGATCACGCCGACCAATATTGTTCTGTTAGCGATGGCCGTATGCGACATGTTAGCTCTTACCTTTCCCGCGCCTTGGTTTTTCTATCTGTACACGATGGGCAATTAC
 H  M  I  T  P  T  N  I  V  L  L  A  M  A  V  C  D  M  L  A  L  T  F  P  A  P  W  F  F  Y  L  Y  T  M  G  N  Y

CACTTAGCTCTTGGTCCATCCTCCTTATGTATTGCTTACCAAACCATGTGCGAAGTGCTTCCTAATGCGTTTCGCACAACTTCTATTTGGTTAACAGTGTTACTCGCGGTA
 H  L  A  L  G  P  S  S  L  C  I  A  Y  Q  T  M  C  E  V  L  P  N  A  F  R  T  T  S  I  W  L  T  V  L  L  A  V

CAACGATACATTTACATCTGTCACCCCACAACGGCTCGAACCTTGTGCACAGTTCCGAACGTGTCGAAGACGATTGCGCTTATTACAGCTCTGGCCTTTGGCCATATGATA
 Q  R  Y  I  Y  I  C  H  P  T  T  A  R  T  L  C  T  V  P  N  V  S  K  T  I  A  L  I  T  A  L  A  F  G  H  M  I

CCAAAGTTTTTTGATGAGATCTTCTACGACGTCGTACTACCGATGGACGGCAAGGCGACAACGGTGTGTGCTCGGAAGATGGCACGTTGGTTAATTCTCTTCACCCCAGAT
 P  K  F  F  D  E  I  F  Y  D  V  V  L  P  M  D  G  K  A  T  T  V  C  A  R  K  M  A  R  W  L  I  L  F  T  P  D

GTCTTCTACAACGTGTACTATACGTTCAAGATGGTATTCGTCAACCTGGTTCCCTGTATCGTGTTGGTTGCCCTTAATATTCTATTGTTCCGTGCCCTTCGCCGAGCCAAG
 V  F  Y  N  V  Y  Y  T  F  K  M  V  F  V  N  L  V  P  C  I  V  L  V  A  L  N  I  L  L  F  R  A  L  R  R  A  K

GAAAACAAGAAGAAGCTACAGGAGAAAAACAAACAGGCAAGCAAACGAAGTCGTGATTCTCAAAGCACAACACTGATGCTAATTGTCGTCGTCACCGTGTTCCTTGCTACC
 E  N  K  K  K  L  Q  E  K  N  K  Q  A  S  K  R  S  R  D  S  Q  S  T  T  L  M  L  I  V  V  V  T  V  F  L  A  T

GAAATCCCTTTGACAAGCGTTCTATTCGTGCATGCGTTACATAACAATGCGGTGATCTCTGTGGACTACGGCGTCTTGAATACTGTGGTAACCTTTGTTAACTTCTCATT
 E  I  P  L  T  S  V  L  F  V  H  A  L  H  N  N  A  V  I  S  V  D  Y  G  V  L  N  T  V  V  T  F  V  N  F  S  L

AATGATGAGTTACCCCCTCAACTTCGCCATCTATTGCGGGATGAGCAAACAGTTCCGGAACACTTTTCGGGAGCTCTTCATCACGGGCTGGTTGCGTGGATCCGGGGAGCT
  M  M  S  Y  P  L  N  F  A  I  Y  C  G  M  S  K  Q  F  R  N  T  F  R  E  L  F  I  T  G  W  L  R  G  S  G  E  L

TAGTAAAGCCCAGTTGATGAACACGCTCGAAACAACCATATAGATCAAGCCTTGCGATTTCCCCGTTCCAAGCAAAAGAAAAAACGATGTCTTCAGGCCAAAAGCTGTTCA
  S  K  A  Q  L  M  N  T  L  E  T  T  I  *
TATGGGTTTCATTCGCAGTCTATCCTAGCCTTGTAACACGAATATTACTTAACTAAGTTGCAGGAAATGATTTCACTGACCAACATCGTAAGAAACGTCAAAAGGTCAGTA
CCCACTGGACCTGGGAATACTTGATTTCACTTTTTTATCAGCTTGCCACTATAATTATTAGTAAACAATGTAGTAGAAAGAGTTACACTTTGGCTAATGAAATTCTGGTTT
CTGTGATCTGTGCCAAAAATACTTATTTTTTCTTCTATCTATTTCAAAGAACTAAAACGTTATCGCTGTTATCTGAAGGTATGTGATGTGAAATCAGATATCACAACGGTG
CAGCCTCTATACTCCCATAAGGAAAATGGGGACCGTTATTGGTTAGTAACGCGCGTAGATGCGGCATGATTAGTGTTTTCTTCAGAAATTATGTGAGGACAATAACTGTGC
TCTTCCTTGTTCTGTGGGCCTGTAAAGTAGCTCCTGGCGATTTCCTAAATTAAATAAACAAAGTACGGATTATGTTTGTTGTTGCAAGACGATATGCTGAGTCTTAGCTGA
TCAACATTCATTCCGTAACTGTCAGGCTTTAGGACGTTGTTTGCAACATGCTTACGGATTTGCAAAAAAAATGCTATCGTTTACCCGTGCTTTGTTATGAAATTTTTGATGA
AGCGGTGCTGTACGGTCACGGCACGATCTAAGTAGTTCGGTATGGACATAATTAGGCGATACGCTCTCAAATATTTGTGGAAATCAATCGTCGTTTTTCTTTCCTAGCCCC
CTAATCCGACCGGCGAAACAATGTAAATCGTCTAGCCGACCTGCGAAGGGTACGTCCAACGACACTTCACCCGCTGGTATTGGTAAATACCGCTTCCCTCTTCATGCGCTG
TCCTGTTAACCGCTGTGAGCCGCAAAGGTTAGATACTGTTGGATTATGTGACAGGTAGATTTCCTGCAAGTCTTCATGATACGTTATCGTATTACGATGCATACAGTTGAA
AGTAACAGCGTGAAAATACTGATATGAAATTAGTTCTGTGGACTAGAAGGCTCTTGTTTTGGTCATGCCATACTACAAATGCAAATTACATTAGACCAATATAAACTAAAA
TAGAAAAAAGGGAATCACCAACTATCAAAATAAACAGTAATATTTTTTAAAAACAA
```

B) Multiple sequence alignment:

```
D. melanogaster  ------------------------------------------------------------MDNYT
```


A. mellifera
I. scapularis
T. urticae
V. destructor
V. jacobsoni
D. melanogaster

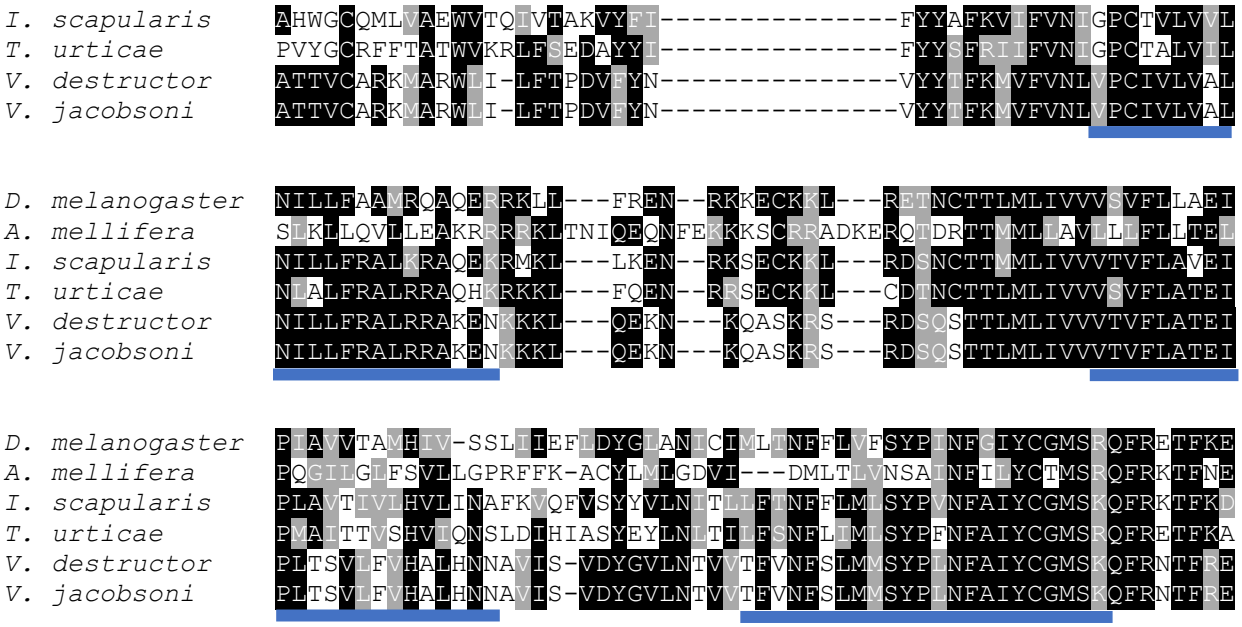
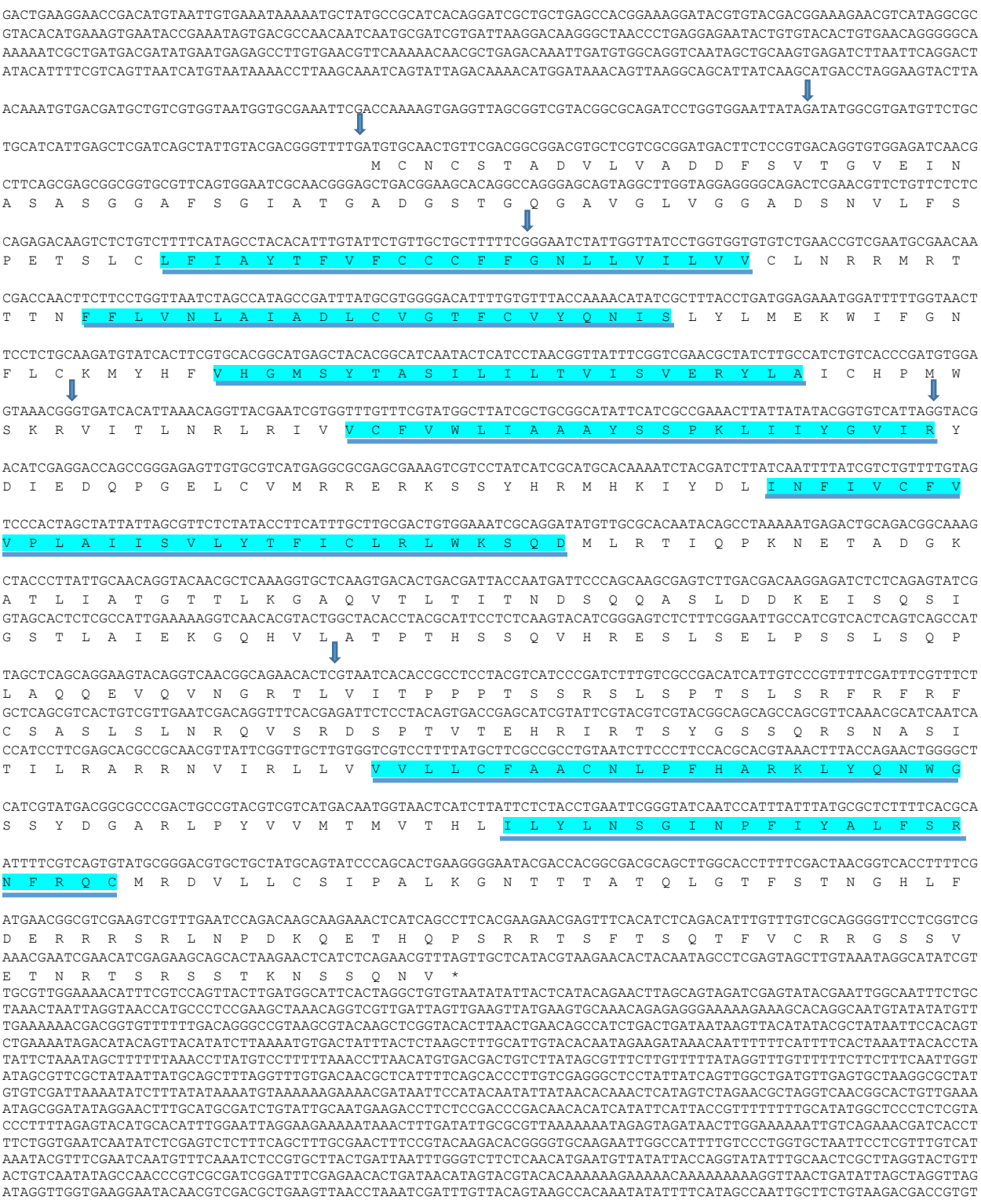


Figure 3.19 A) Gene structure for Ast B receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of All B receptor from various orthologous species across Phylum Arthropoda

Trissin receptor:

A) Gene structure

V. destructor:



```
CAGCGGGCATCTGTTCTTCAAAGTGTGCTTCGCGCGCTTGACGTCAAGCTCATGATACAGAATCGAAACTTACTTGTTGTTATTGTTGTTGTTGTTGTTAAGTTATTTTCA
AGCATTTGAAGTCGAAAAGTATTGGATAATCGAATCAGCATTGTAACTGATCATTTTTTAATGCACAGCCTCGTCTGCGGGTTAATAATCACGCCTCGAAATAGCTTATCT
GAGCGACCCAGGTAACTTTGTGCTTGTATCGTGCAGACATCTACATCCGTCTTCATGTTTACCCGATCGGTAGTAGGCAAGAGAATCCACTGGAAGTGAATTCGACGCTAG
CGCTAGAGGTCGCTTCGGATGCCGATCTAACTAACCGACAATATCAACTGGCCCGCTACGACTATTTGATGGGTCATGTCAGTCTGCTAGCCGGTTACCAGCGTCGCGCCT
GTCTTACAATGTCTGATCGTTCTATCTGGGCATACGCGTCGGCCTACGGGTTTCAGCTAAGGTACATGTCAAAAAGTAGGAACATTCCCAATACGAATGATATCAAGCTAA
CGAAGAGACTTGACGATGAAAGCACGCTTGGTAGTGGGTTCGTGGTACAACATCGATTATAGTCACTTAGAGCTCAAGGTCTCGGTTGATGACCAAACCGGGACGCTATTC
ATGTCAAGTGGAGTCGACGATCGACAGATCTTTCTACTCGGCTACCCCAACAACTACAATTCGGCCCCCAATTCATCAGCTGAGCGCTAACAATTGATCTCCTCGATATGG
GGCTATCCTGCTTCATGGGCACATTGCGTTGTATATATGATGCCTACGAGGCCCTGTAAAAGTGTACAGATTATCGAATAACGAACGAAGGCAAATACCTGGCAGTGTAAA
TAGTTACTTTGTTCTATTAATGACTTGTACCAGCTGGCTAGACTAAGGTATCCTATGAATGTAGCTAGCAAGTCAAATGAATGGCTTTCAACTCTATTCAATAACATTGCT
CGGTGGCCATTTAAGCTTGGGTTGCAGCTGCGGATGAGTGTTTTAATTTTTTTTACTTATTTTAGACAGCATATTGAATTAAATTTGTGTGCATTTGTCTACGGTCACCAA
TCATTACTTTGTTTATTCCCTTCTGTCGTCTGCTACATGCAAACACTTAATAATTTCTTTTTTTTTTTCCTGTCACATGGTAGAGGACCTTCACAGTAAATTTACCTGATC
AGAAAACGCTTGCAAATCGTACCACTAAACAACACCTCAGCGTCAAACACGCGACGATTTTTTTCTCGACCTCATTTGTACTTCGAAACTTAACTGATACCTTTATTTGAA
TGCGTTCTCATAAACATCGAATGGCATCCATAATCAAGGAACTATTCCCTTCAAATCGACGCATAACACACATCTAGGTCTTGCAGTGTTTCAAGCTGATTTTATC
AGTTGGTAGTATTTACCACAGGACTCGCACACGTTTGCGGTTACCATTCACGTGTAAAGCTCCACAGGATTTTGATGATGTGACAATGAAGATCTGTGATGTAAATAGAAA
TCACTACCGATTAATAAAATAAGGTACGTTGCTGTCAATGTTAAGTAGACTCGTTGTTTATGAGTACGTTTACAGCGCGATAGGAACAAGGATTTTTTCGACGACATGTCT
CCCGTACAGATAGCTTACTTTACAAATGACCTTTTATACCGCCCTAGAGAACTGAATTGTATCTTCCTTTAGACAAAATCGATTATTTTTGATAATCCCTTGTAAAATGTA
TACAACATTCTCGGTTGTTAGTATACCATATAGCTTGCACTGGTCTAGCTCTTCTAAGACCTCTACCCCGCAACGAAGCTATACCGATCAGTTCGCAGTGTATTGGCAGTG
TACATGTGCTATATACTCAGTGCAATATTACCTTCCTTTCATTATACAACTCCATATGCTGAAACGCAAATAAATCGAACACAGTCAACAGGCACCCCTACACGGACATCA
CATCTTCAAATGTATAGATACTTGTGTGCGCTGAGCGAAGTCTTTAGGTACATCTGTAATTGGTAAAGTTGTAAACATGTAAAGCACTAGAGCTAATGTAAATA
```

V. jacobsoni:

```
TGTCGTGGTAATGGTGCGAAATTCGACCAAAAGTGAGGTTAGCGGTCGTACGGCGCAGATCCTGGTGGAATTATAGATATGGCGTGATGTTCTGCTGCATCATTGAGCTCC

ATCAGCTATTGTACGACGGGTTTTGATGTGCAACTGTTCGACGGCGGACGTGCTCGTCGCGGATGACTTCTCCGTGACAGGTGTGGAGATCAACGCTTCAGCGAGCGGCGG
                          M  C  N  C  S  T  A  D  V  L  V  A  D  D  F  S  V  T  G  V  E  I  N  A  S  A  S  G  G
TGCGTTCAGTGGAATCGCAACGGGAGCTGACGGAAGCACAGGCCAGGGAGCAGTAGGCTTGGTAGGAGGGGCAGACTCGAACGTTCTGTTCTCTCCAGAGACAAGTCTCTG
  A  F  S  G  I  A  T  G  A  D  G  S  T  G  Q  G  A  V  G  L  V  G  G  A  D  S  N  V  L  F  S  P  E  T  S  L  C

TCTTTTCATAGCCTACACATTTGTATTCTGTTGCTGCTTTTTCGGGAATCTATTGGTTATCCTGGTGGTGTGTCTGAACCGTCGAATGCGAACAACGACCAACTTCTTCCT
  L  F  I  A  Y  T  F  V  F  C  C  C  F  F  G  N  L  L  V  I  L  V  V  C  L  N  R  R  M  R  T  T  T  N  F  F  L

GGTTAATCTAGCCATAGCCGATTTATGCGTGGGGACATTTTGTGTTTACCAAAACATATCGCTTTACCTGATGGAGAAATGGATTTTTGGTAACTTCCTCTGCAAGATGTA
  V  N  L  A  I  A  D  L  C  V  G  T  F  C  V  Y  Q  N  I  S  L  Y  L  M  E  K  W  I  F  G  N  F  L  C  K  M  Y

TCACTTCGTGCACGGCATGAGCTACACGGCATCAATACTCATCCTAACGGTTATTTCGGTCGAACGCTATCTTGCCATCTGTCACCCGATGTGGAGTAAACGGGTGATCAC
  H  F  V  H  G  M  S  Y  T  A  S  I  L  I  L  T  V  I  S  V  E  R  Y  L  A  I  C  H  P  M  W  S  K  R  V  I  T

ATTAAACAGGTTACGAATCGTGGTTTGTTTCGTATGGCTTATAGCTGCGGCATATTCATCGCCGAAACTTATTATATACGGTGTCATTAGGTACGACATCGAGGACCAGCC
  L  N  R  L  R  I  V  V  C  F  V  W  L  I  A  A  A  Y  S  S  P  K  L  I  I  Y  G  V  I  R  Y  D  I  E  D  Q  P

GGGAGAGTTGTGCGTCATGAGGCGCGAGCGAAAGTCGTCCTATCATCGCATGCACAAAATCTACGATCTTATCAATTTTATCGTCTGTTTTGTAGTCCCACTAGCTATTAT
  G  E  L  C  V  M  R  R  E  R  K  S  S  Y  H  R  M  H  K  I  Y  D  L  I  N  F  I  V  C  F  V  V  P  L  A  I  I

TAGCGTTCTCTATACCTTCATTTGCTTGCGACTGTGGAAATCGCAGGATATGTTGCGCACAATACAGCCTAAAAATGAGACTGCAGACGGCAAAGCTACCCTTATTGCAAC
  S  V  L  Y  T  F  I  C  L  R  L  W  K  S  Q  D  M  L  R  T  I  Q  P  K  N  E  T  A  D  G  K  A  T  L  I  A  T

AGGTACAACGCTCAAAGGTGCTCAAGTGACACTGACGATTACCAATGATTCCCAGCAAGCGAGTCTTGACGACAAGGAGATCTCTCAGAGTATCGGTAGCACTCTCGCCAT
  G  T  T  L  K  G  A  Q  V  T  L  T  I  T  N  D  S  Q  Q  A  S  L  D  D  K  E  I  S  Q  S  I  G  S  T  L  A  I
TGAAAAAGGTCAACACGTACTGGCTACACCTACGCATTCCTCTCAAGTACATCGGGAGTCTCTTTCGGAATTGCCATCGTCACTCAGTCAGCCATTAGCTCAGCAGGAAGT
  E  K  G  Q  H  V  L  A  T  P  T  H  S  S  Q  V  H  R  E  S  L  S  E  L  P  S  S  L  S  Q  P  L  A  Q  Q  E  V

ACAGGTCAACGGCAGAACACTCGTAATCACACCGCCTCCTACGTCATCCCGATCTTTGTCGCCGACATCATTGTCCCGTTTTCGATTTCGTTTCTGCTCAGCGTCACTGTC
  Q  V  N  G  R  T  L  V  I  T  P  P  P  T  S  S  R  S  L  S  P  T  S  L  S  R  F  R  F  R  F  C  S  A  S  L  S

GTTGAATCGACAGGTTTCACGAGATTCTCCTACAGTGACCGAGCATCGTATTCGTACGTCGTACGGCAGCAGCCAGCGTTCAAACGCATCAATCACCATCCTTCGAGCACG
  L  N  R  Q  V  S  R  D  S  P  T  V  T  E  H  R  I  R  T  S  Y  G  S  S  Q  R  S  N  A  S  I  T  I  L  R  A  R
CCGCAACGTTATTCGGTTGCTTGTGGTCGTCCTTTTATGCTTCGCCGCCTGTAATCTTCCCTTCCACGCACGTAAACTTTACCAGAACTGGGGCTCATCGTATGATGGCGC
  R  N  V  I  R  L  L  V  V  V  L  L  C  F  A  A  C  N  L  P  F  H  A  R  K  L  Y  Q  N  W  G  S  S  Y  D  G  A

CCGACTGCCGTACGTCGTCATGACAATGGTAACTCATCTTATTCTCTACCTGAATTCGGGTATCAATCCATTTATTTATGCGCTCTTTTCACGCAATTTTCGTCAGTGTAT
  R  L  P  Y  V  V  V  M  T  M  V  T  H  L  I  L  Y  L  N  S  G  I  N  P  F  I  Y  A  L  F  S  R  N  F  R  Q  C  M

GCGGGACGTGCTGCTATGCAGTATCCCAGCACTGAAGGGGAATACGACCACGGCGACGCAGCTTGGCACCTTTTCGACTAACGGTCACCTTTTCGATGAACGGCGTCGAAG
  R  D  V  L  L  C  S  I  P  A  L  K  G  N  T  T  T  A  T  Q  L  G  T  F  S  T  N  G  H  L  F  D  E  R  R  R  S
TCGTCTGAATCCAGACAAGCAAGAAACTCATCAGCCTTCACGAAGAACGAGTTTCACATCTCAGACATTTGTTTGTCGCAGGGGTTCCTCGGTCGAAACGAATCGAACATC
  R  L  N  P  D  K  Q  E  T  H  Q  P  S  R  R  T  S  F  T  S  Q  T  F  V  C  R  R  G  S  S  V  E  T  N  R  T  S
GAGAAGCAGCACTAAGAACTCATCTCAGAACGTT TAG TTGCTCATACGTAAGAACACTACAATAGCCTCGAGTAGCTTGTAAATAGGCATATCGTCGCGTTGGAAAACATT
  R  S  S  T  K  N  S  S  Q  N  V  *
TCGTCCAGTTACTTGATGGCATTCACTAGGCTGTGTAATATATTACTCATACAGAACTTAGCAGTAGATCGAGTATACGAATTGGCAATTTCTGCTAAACTAATTAGGTAA
CCATGCCCTCCGAAGCTAAACAGGTCGTTGATTAGTTGAAGTTATGAAGTGCAAACAGAGAGGGAAAAAGAAAGCACAGGCAATGTATATATGTTTGAAAAAACGACGGTG
TTTTTTGACAGGGCCGTAAGCGTACAAGCTCGGTACACTTAACTGAACAGCCATCTGACTGATAATAAGTTACAGATACGCTATAATTCCACAGTCTGAAAATAGACATAC
```

```
AGTTACATATCTTAAAATGTGACTATTTACTCTAAGCTTTGCATTGTACACAATAGAAGATAAACAATTTTTTCATTTTCACTAAATTACACCTATATTCTAAATAGCTTT
TTTAAACCTTATGTCCTTTTTAAACCTTAACATGTGACGACTGTCTTATAGCATTTCTTGTTTTTATAGGTTTGTTTTTTCTTCTTTCAATTGGTATAGCGTTCGCTATAA
TTATGCAGCTTTAGGTTTGTGACAACGCTCATTTTCGGCACCCTTGTCGAGGGCTCCTATTATCAGTTGGCTGATGTTGAGTGCTAAGGCGCTATGTGTCGATTAAAATAT
CTTTATATAAAATGTAAAAAAGAAAACGATAATTCCATACAATATTATAACACAAACTCATAGTCTAGAACGCTAGGTCAACGGCACTGTTGAAAATAGCGGATATAGGAA
CTTTGCATGCGATCTGTATTGCAATGAAGACCTTCTCCGACCCGACAACACATCATATTCATTACCGTTTTTTTTTGCATATGGCTCCCTCTCGTACCCTTTTAGAGTACA
TGCACATTTGGAATTAGGAAGAAAAATAAACTTTGATATTGGGCGTTAAAAAAATAGAGTAGATAACTTGGAAAAAATTGTCAGAAACGATCACCTTTCTGGTGAATCAAT
ATCTCGAGTCTCTTTCAGCTTTGCGAACTTTCCGTACAAGACACGGGGTGCAAGAATTGGCCATTTTGTCCCTGGTGCTAATTCCTCGTTTGTCATAAATACGTTTCGAAT
CAATGTTTCAAATCTCCGTGCTTACTGATTAATTTGGGTCTTCTCAACATGAATGTTATATTACCAGGTATATTTGCAACTCGCTTAGGTACTGTTACTGTCAATATAGCC
AACCTGTCGCGATCGGATTTCGAGAACACTGATAACATAGTACGTACACAAAAAAGAAAAACAAAAAAAAAAGGTTAACTGATATTAGCTAGGTTAGATAGGTTGGTGAAG
GAATACAACGTCGACGCTGAAGTTAACCTAAATCGATTTGTTACAGTAAGCCACAAATATATTTTCATAGCCAATTGCTTCTGTAAGACGACCGTGTCAGCGGGAATCTGT
TCTTCAAAGTGTGCTTCGCGCGCTTGACGTCAAGCTCATGATACAGAATCGAAACTTACTTGTTGTTATTGTTGTTGTTGTTGTTAAGTTATTTTCAAGCATTTGAAGTCG
AAAAGTATTGGATAATCGAATCAGCATTGTAACTGATCATTTTTTAATGCACAGCCTCGTCTGCGGGTTAATAATCACGCCTCGAAATAGCTTATCTGAGCGACCCAGGTA
ACTTTGTGCTTGTATCGTGCAGACATCTACATCCGTCTTCATGTTTACCCGATCGGTAGTAGGCAAGAGAATCCACTGGAAGTGAATTCGACGCTAGCGCTAGAGGTCGCT
TCGGATGCCGATCTAACTAACCGACAATATCAACTGGCCCGCTACGACTATTTGATGGGTCATGTCAGTCTGCTAGCCGGTTACCAGCGTCGCGCCTGTCTTACAATGTCT
GATCGTTCTATCTGGGCATACGTGTCGGCCTACGGGTTTCAGCTAAGGTACATGTCAAAAAGTAGGAACATTCCCAATACGAATGATATCAAGCTAACGAAGGGACTTGAC
GATGAAAGCACGCTTGGTAGTGGGTTCGTGGTACAACATCGATTATAGTCACTCAGAGCTGAAGGTCTCGGTTGATGACCAAACCGGGACGCTATTCATGTCAAGTGGAGT
CGACGATCGACAGATCTTTCTACTCGGCTACCCCAACAACTACAATTCGGCCCCCAGTTCATCAGCTGAGCGCTAACAATTGGTCTCCTCGATATGGGGCTATCCTGCTTC
ATAGGCACATTGCGTTGTATATATGATGCCTACGAGGCCCTGTAAAAGTGTACAGATTATCGAATAACGAACGAAGGCAAATACCTGGCAGTGTAAATAGTTACTTTGTTG
TATTAATGACTTGTACCAGCTGGCTAGACTAAGGTATTCTATGAATGTAGCTAGCAAGTCAAATGAATGGCTTTCAACTCTATTCAATAACATTGCTCGGTGGCCATTTAA
GCTTGGGTTGCAGCTGCGGATGAGTGTTTTAATTTTTTTTACTTATTTTAGACAGCATATTGAATTAAATTTGTGTGCATTTGTCTACGGTCACCAATCATTACTTTGTTT
ATTCCCTTCTGTCGTCTGCTACATGCAAACACTTAATAATTTCTTTTTTTTTTCCTGTCACATGGTAGAGGACCTTCACAGTAAATTTACCTGATCAGAAAACGCTTGCAA
ATCGTACCACTAAACAACACCTTAGCGTCAAACACGCGACGATTTTTTTCTCGACCTCATTTGTACTTCGAAACTTAACTGATACCTTTATTTGAATGCGTTCTCATAAAC
ATCGAATGGCATCCATAATCAAGGAACTATTCCCTTCAAATCAGACTGACGCATAACACACATCTAGGTCTTGCAGTGTTTCAAGCTGATTTTATCAGTTGGTAGTATTTA
CCACAGGACTCGCACACGTTTGCGGTTACCATTCACGTGTAAAGCTCCACAGGATTTTGATGATGTGACAATGAAGATCTGTGATGTAAATAGAAATCACTACCGATTAAT
AAAATAAGGTACGTTGCTGTCAATGTTAAGTAGACTCGTTGTTTATGAGTACGTTTACAGCGCGATAGGAACAAGGATTTTTTCGACGACATGTCTCCCGTACAGATAGCT
TACTTTACAAATGACCTTTTATACCGCCCTAGAGAACTGAATTGTATCTTCCTTTAGACAAAATCGATTATTTTTGATAATCCCTTGTAAAATGTATACAACATTCTCGGT
TGTTAGTATACCATATAGCTTGCACTGGTCTAGCTCTTCTAAGACCTCTACCCCGCAACGAAGCTATACCGATCAGTTCGCAGTGTATTGGCAGTGTACATGTGCTATATA
CTCAGTGCAATATTACCTTCCTTTCATTATACAACTCCATATGCAAATAAATCGAACACAGTCAACAGGCACCCCTACACGGACATCACATCTTCAAATGTATAGATACTT
GTGTGCGCTGAGCGAAGTCTTTAGGTACATCTGTAATTGGTAAAGTTGTAAACATGTAAAGCACTAGAGCTAATGTAAATATCGGATCGATGATGAAAATAAAATACCTAA
ATTTATCACATAA
```

(B) Multiple sequence alignment:

```
D. melanogaster  MIMTMMQTVRAWQQESDVEHRKQHKQRWRPDGAHISAAYDLNSDNDDGHHRVVHNQNNGS
I. scapularis    ------------------------------------------------------------
T. urticae       ------------------------------------------------------------
V. destructor    ------------------------------------------------------------
V. jacobsoni     ------------------------------------------------------------

D. melanogaster  PNSSPNQSTSAFRQRQPHHPPTGQQPPRLPCTVTHFSAHWKTLLILLTLLSASTLTASAN
I. scapularis    ------------------------------------------------------------
T. urticae       ------------------------------------------------------------
V. destructor    ------------------------------------------------------MCN
V. jacobsoni     ------------------------------------------------------MCN

D. melanogaster  VTSTISPPINGSSTDYILLYGESTTSLVPALTTGLSGDGSGAVIEDEEDAEKASEYIFDR
I. scapularis    ---MDALSANDSNANLLDNFGNGSGLWIPG---GAEED----------------EDVFQR
T. urticae       ------MLLDNNETFLSLMIIDNDTMRMNESYQGED------------------DFLLTN
V. destructor    CSTADVLVADDFSVTGVEINASASGGAFSGIATGADGSTGQGAVGLVGGAD---SNVLFS
V. jacobsoni     CSTADVLVADDFSVTGVEINASASGGAFSGIATGADGSTGQGAVGLVGGAD---SNVLFS

D. melanogaster  TDVRIIFITLYTLVFCCCFFGNLLVILVVTLSRRLRSITNFFLANLAFADFCVGLFCVMQ
I. scapularis    SDVRVGLIFTYSLVFCCCFFGNLLVVVVVVHRRMRTITNFFLTNLAVADLCVGLFCVYQ
T. urticae       KSFKIGLIIFYILIFFSTFVGNLVVILVVILQKKMRNSTNFFFTNLAFSDLCVGLFCIFQ
V. destructor    PETSLCLFIAYTFVFCCCFFGNLLVILVVCLNRRMRTTTNFFLVNLAIADLCVGTFCVYQ
V. jacobsoni     PETSLCLFIAYTFVFCCCFFGNLLVILVVCLNRRMRTTTNFFLVNLAIADLCVGTFCVYQ

D. melanogaster  NLSIYLIESWVFGEFLCRMYQFVHSLSYTASIFILVVICMERYFAIVHPITCKQILTAAR
I. scapularis    NLSIYLMDEWPLGDFLCRMYFFVQALSYTASVGILTVICVERYLAIVHPMWSKHVITTRR
T. urticae       DLSSFLQPTWSFGVIMCKMYHFTQTMSYTASIFTMVIISIERYLAICYPIRAKKVLQMRN
V. destructor    NISLYLMEKWTFGNFLCKMYHFVHGMSYTASILILTVISVERYLAICHPMWSKRVITLNR
V. jacobsoni     NISLYLMEKWTFGNFLCKMYHFVHGMSYTASILILTVISVERYLAICHPMWSKRVITLNR
```

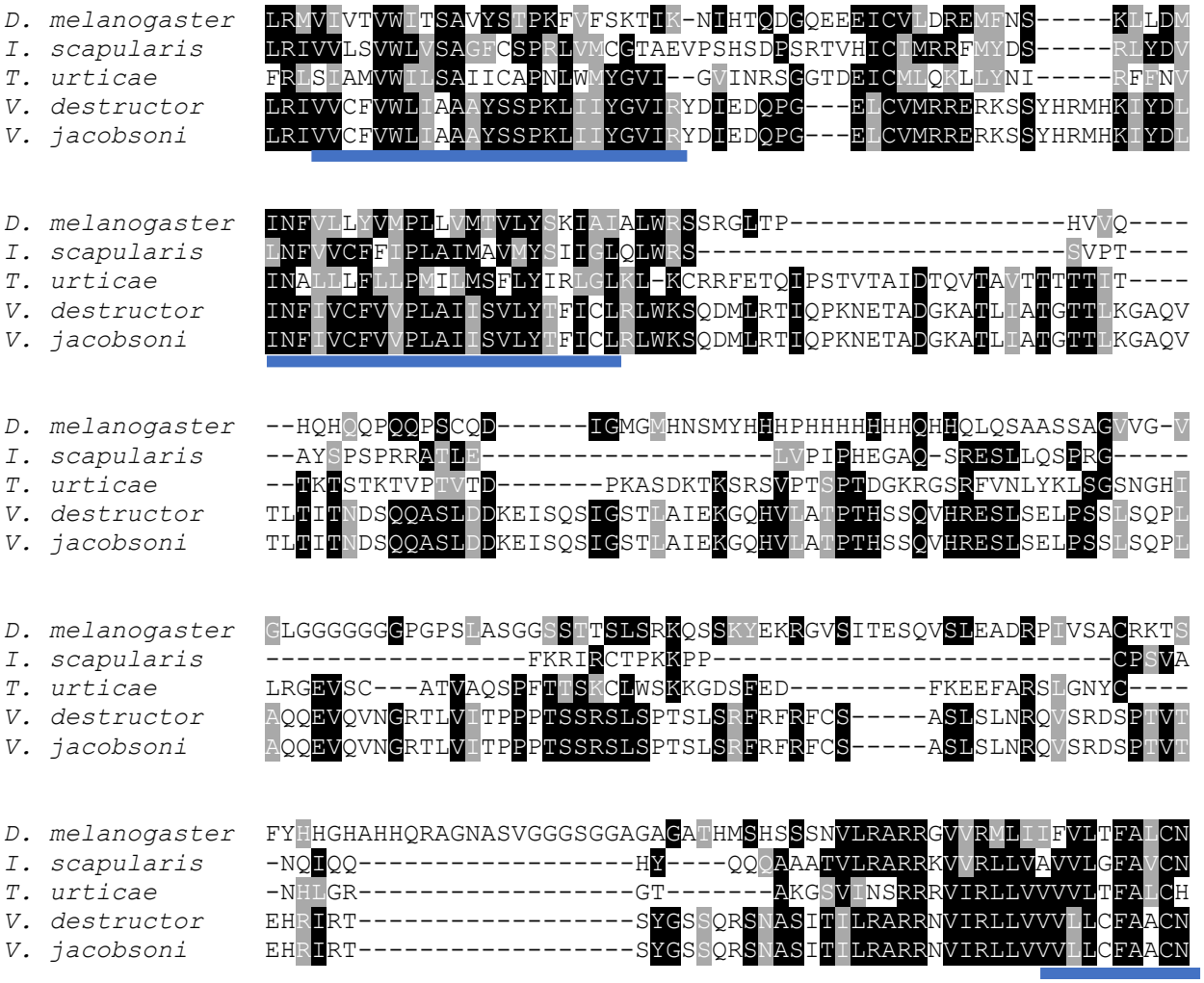


Figure 3.20 A) Gene structure for trissin receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of trissin B receptor from various orthologous species across Phylum Arthropoda

EFLamide receptor:

A) Gene structure

V. destructor EFLR1

```
AATAGCGGCGAACAACGAGCAACACCGAAATTAAAGACATCAATAAATGTGTAACCTCGAACAAATTAACAAGCGACGATTGAGTGTGTTAACAACTTGGTGGCGGACGCT
GTCGACACCGTTACTATTTAACAGTTTTCATTATTCGGTTCCGATAAAGTACCGTTTCACAGAAATCATTGTTAGACGCTTTCTCGTTGTCCTCTCTGTGTGAAAGCAATA
GCAACGAGGAAAAAGAATAAAAGACCTAGGATTATAGACTACACAGTGGTCGAGCGCTGCTTCCCATGTGGAGCAGTTGTCTTGACCGAAAACGAGCGGTGTTTTGTCAAA
GACACTTAGCAGTTAACTCACGACCCATGAAAACTTTGCCGCTAGCCAAACGTTCCGCCTATCTGTCCAGTCGGTCATAGGCAGTCGCACCAACCATGAAGTGGCCCATCG
                                                                                                  M  K  W  P  I
TCGCATCTACCTGGCTCAGTGTTGCTGTTGCCAACGTGGTACCAGTTGTGGACCCATTGTGTTCGACGAATATTGCCAACAACAACAGGACCAGCATACAAATGGCAATTG
V  A  S  T  W  L  S  V  A  V  A  N  V  V  P  V  V  D  P  L  C  S  T  N  I  A  N  N  N  R  T  S  I  Q  M  A  I

ATTGTGACGACAGCGACTTGTACGACGCTATCAATGGAGGTTCCGTAACGAATTCGAACACCTCTATCGGTTCCTGGTCCTCTTCAGCCTCATCGATGGCCTCGTCGTCGG
D  C  D  D  S  D  L  Y  D  A  I  N  G  G  S  V  T  N  S  N  T  S  I  G  S  W  S  S  S  A  S  S  M  A  S  S  S
CGGCGTCGAATGGGGTGCTGACAGGCAGCCAGAGTGTCAATGCCACGAGCTCTGTTAGGGCGAATTCACTGACGAGGGAGCCGAGGTATCGTCCCTTGGTAGACACCATCG
A  A  S  N  G  V  L  T  G  S  Q  S  V  N  A  T  S  S  V  R  A  N  S  L  T  R  E  P  R  Y  R  P  L  V  D  T  I
TTGGGACGGCGGTTCTTGGCGCCATCCTTCTTATCGGAGTTGCAGGCAATATAATGGTTGTTGCAGTGGTCGCCCAAACGCGCTCAATGCGAACTCCGACGAACTGCTATC
V  G  T  A  V  L  G  A  I  L  L  I  G  V  A  G  N  I  M  V  V  A  V  V  A  Q  T  R  S  M  R  T  P  T  N  C  Y

TCGTCTCGTTGTCAGTAGCCGACCTCATGGTTCTGCTATCGGCAATCCCGAATGAGATCCTCGCCCAGTACATCCTTGAGGACGAATGGATCTTTGGCCGAGCTGGCTGTG
L  V  S  L  S  V  A  D  L  M  V  L  L  S  A  I  P  N  E  I  L  A  Q  Y  I  L  E  D  E  W  I  F  G  R  A  G  C

CAATGTTTGTTTTTTTCCAGTACCTGGGCATTAACGCCTCGTCTTTATCAATTACGGCATTTACCGTCGAGCGATATATCGCCATCTGCCTGCCAATGAAGGCGCAGACCA
A  M  F  V  F  F  Q  Y  L  G  I  N  A  S  S  L  S  I  T  A  F  T  V  E  R  Y  I  A  I  C  L  P  M  K  A  Q  T

TGTGCACGGTGAAACGTGCTAAGAAGATTATCCTCGCTGTGTGGTTGTTTGCGCTGTGTTACTGTTGCCCATGGTTATTTTTTCTTACGACGACACGGCCAATCTACTACC
M  C  T  V  K  R  A  K  K  I  I  L  A  V  W  L  F  A  L  C  Y  C  C  P  W  L  F  F  L  T  T  T  R  P  I  Y  Y

GGGGCCAGGAGCACCGTTACGTGGAGACGTGCACGTATTCACTGCACCGATCGTTCTATATGCGTTTTTTTTTTGCCGACATCATTCTATTCTATGTAATTCCGCTCATTC
R  G  Q  E  H  R  Y  V  E  T  C  T  Y  S  L  H  R  S  F  Y  M  R  F  F  F  A  D  I  I  L  F  Y  V  I  P  L  I

TATCATGTGTCTTGTACGGGCGAATGGCCCAAGTGTTATTCCATACGGACATTGGTGCGATGACGACGGCAGCGAAGACTGCCAACGGGCAGACTACAGCTAAAGTCAATG
L  S  C  V  L  Y  G  R  M  A  Q  V  L  F  H  T  D  I  G  A  M  T  T  A  A  K  T  A  N  G  Q  T  T  A  K  V  N
                                                                                                  ⇓
CTAATATGAACAACCAGTCCAGAATTCAGGTGGTGAAAATGCTGATCGTGGTCGTTGTACTTTTCGCGACGTTGTGGCTACCATATCGTGTGCTGCTCGTCTATAATTCGT
A  N  M  N  N  Q  S  R  I  Q  V  V  K  M  L  I  V  V  V  V  L  F  A  T  L  W  L  P  Y  R  V  L  I  V  Y  N  S

TTGCCAGCAAGCCTTACCTCGAGCTGTGGTACCTGATGCTCTGCAAGACGCTCATCTTCGTCAATTCGGCCATCAATCCTGTGCTCTACAACGCCATGTCGGTCAAGTTCC
F  A  S  K  P  Y  L  E  L  W  Y  L  M  L  C  K  T  L  I  F  V  N  S  A  I  N  P  V  L  Y  N  A  M  S  V  K  F
                                                                                                  ⇓
GGCGGGCCTTTCGGCGGACCTTATCCTGCGATTCTAAAAATCGTTTCGAAAATCGCACGTTTACTACAACAATGACAACGGTCAGGCACACGATAAAGCCGAAGGATCAAT
R  R  A  F  R  R  T  L  S  C  D  S  K  N  R  F  E  N  R  T  F  T  T  T  M  T  T  V  R  H  T  I  K  P  K  D  Q
ATTGACGGGACGCGAACTCCTAAATTACTTTTGCCTAGAAGCTCGCAGAAATGAGGCAAACAACCGATTTCGAGAAGTACCTCGTCGTAAATGGCAACACCTCAACGCCGA
Y  *
ACGACACAGAAAATTCGTCGAAAATTGCGCTTGCCGATAAACCTAACCAAGAGAATAAGAATGAATGCAACCTGCAAGCTGCGACTCACCATTGTTAAACGACTTCACATC
CCCCTTAGGCAGCCCTGTTTCGAAGATATTTTGTTGCAGAAAATCGCGATGATAACTCAGTTTTACCCTGCACAACGTTGATAGTATTGCTGACCCAATGCTAAAACAGCT
AAATCAATGAGTTCACAAAAGTCATGGCCGCGATTTTTCAACGTAGGGACCTGTAACTGGTCGTCATTGAATGAGTAAATTGAAAGAAAAACATTAAATTGGAATTTTCCT
GACTGAGTATAATTATAGTAGTATCGATAAAGTGAATTAATGTACACTAAATATGAATCTATCGTATACATTTTGGTGGAATGTAGCACTTAAATGTGCTTTGAGCACAAA
AAAAAAACAAGCAATAGATATCAATGCCTGATATCCTAAAATTATAGTTTGTACCGTTGATGCAATAGACCATAACTGCATTGGCCATTGGGTCGTTGTGCTTCTGCAGAC
TGCAGAATCGTCCAACGTGTGCAGATTCTCTGATTCGCTTCACTGACGTATTGCGTTTTGCCAGCCGCCAAGGTGCACTCAAACAAAGTCTACCATAAGTCGACTTCTGCG
TATTTATAATCATATACCCTCACATCGTGTACTTTAACATCCTGTTTGTGTCGTACCAACAAACAAATCAACTTCCTACACGTTTCGCGTAGCTTCCGCCAATAGCCCAGC
CCGCGCAATGAAATCAGTCAATAAAACAAAACCCGAATGCGCGAGATGCACCCAAGACGAATCGAGAGTCTTTTATCTGCAGTTAGAAAGGCAAACGAAGACTCTTTTGTT
TGCTTCCGACGTGAGAATTCCCGTACCACGATATTATTCGTAAGATGAGTGGATCTGGGAGGAGTCGGGGAAAATGTTAACCAGCGGGAATAACTGAACGAGACCGAAGTA
AAATCGAATTGAAGTTAGACCGAGTTTATTTTGCTTCCAGTAATCAGTACTACTATTGCTTATGCTAATAGAATACTAAAGGAAATAGATTCAGCCTCTCTTAGTTGTTTC
GTGAGAACGAATTTTACGTGCAGACGCATAGACCAACACACACACACACACACACACACGTACACGCCCATGCCGGCATGAAAGGATAATAAACAGCATAGAATTATGATA
TCTGTATGGGATGCAACTGCGGGCGTCGTCGGAAAAAAAACAAAAGGTAAGAACAACAATCACATTTTGTAGATTAAGAGCAACACGCCATGTTTTCATGGTATCAATTCAG
ACAAGTATAGAGAGTATAAAAAAAAAATGTGATCTAATTAACATTTTTCACTTGGCCAGGCGGCGAAATGGACACTCAAGGCATGTTAAAGTTACTAACTAAGACAATTTCT
TTCAAAAAGCTAGGCTGTTTACGAAGTCCAGAACTCTTCACAGCGGAAGCAGGTGGGTGGTGTAGCAGCTTGAGGTAGTTCATCTATGCATACACATAAACCACTAGGACC
GTATAACGGACCATTGTAGCGGCAGTACATGCAAATGTAAGATGAATGCAGCCTTAACACTGTCGATTGACACAGACATTTAATTTTAATCTGTATCTTTAAAAATATTTT
CTAGTACAGTTTAGAGTAAAAAAAAATAGTCCTCTTCCTATGGCGGGATCAGATTATCATAAGTAGAGATTCCGAGCAAGGAACATACGACGAATATGTTGAAAACATACAA
TGGATCTCAGTAAATCCGAAGATGACTTTTGCCGTGACAAATGGCTCTATCTATTGTGCATGAGAGCAGAAAACCCGTTGAGGCATTTCCTATTAACTCCGGGAAAGAATA
GGCCGGCTTTGAGCAAGGCACCAACAATCTGTAAACGACACCAGTCTGTCACACAGTTTTGTAAAACGGCTAAAAAAAAAAAAATACGGGGTACATCTACGAGACAGAGTAA
CGATCAGACTCTCTTCTAGAGTTACCCACAACATGTAAAAAAAACTGATTAAAAAAGAGCAGTTTTTCATAAGATATTTTACCTTATTGTACAGTATGATGTTATAAACGA
GGGTTGATGAATATATAAATGACATTGAAAAAAAACAGTCGAACCTGGTGTATATATTTATTCTTCTCACTGAAAAACGCCTCCATATGTAAATGTAATAGAATAATCCGA
GCAGGAGGTAGAGGTGGACAGGGTAAGAAAGGGCAAGAAACGAGAGTGTCCGATGAAAGAGAAATGGTCTCTATGCTGTGGTGTTAATAAAAGCGTCAAAATATTTGTCGG
GTT
```

EFLR 2

```
AGTTCTTAAGAGTTCACCTGTTTCAACGGGCCCAGCCATACGCGTACTCTTCTATCCATTAGCTAACACAGTGCCGTGTCGCCGCTTTTGTGGTTTCGTGCACGCATGTGT
ATGTGACATGGCGATTACCATTGAACTCGATCATTTGTTTGGATGTTTGTTACTCGTGCCCCTGGACGTAACCTAGCAGTCTAACGAGTCGAGAAAGATCTTGTACTGGTT
TTTCAGAAAACAGAAACGAAACCATGTCATATCCAGGCTACGAAACAATTTTAGATGACTACGAAATAATAGCTACAACCATGTCGTCTTCTTCTTTTGATTGTTGCTACA
TGCGCCATGTCTACAATGTGATAGTGTGTCAGCCTGTGCGTTTGTATAATTCACCACTACCCAGCCACGTTTAACTTAGGTCGGCTGAAATTTCAAAGAAAACCGAATCTA
TCGAAGGTCTCGTTTCCGAGTGCTTTCACATTCGATTATACTCGTTAGTTCTGTGCTATTTTCTGTTGTGCCCAGAGCCGTCGCTATCGCATGATGTGCTGCTTCGGAGAA
                                                                                                    M  M  C  C  F  G  E
AAAGTCGCTCCGGCGTTCTGTCACGTAAGACGACGAATGCTGCTACTAGTAACATTATTTCTTGTCATTGATGCGCGAACGGTTAAGTCGACATGCGTCGGCGGAAAAGCA
 K  V  A  P  A  F  C  H  V  R  R  R  M  L  L  L  V  T  L  F  L  V  I  D  A  R  T  V  K  S  T  C  V  G  G  K  A
CATGGCGGCAGCGGAAGCGCAGGCGCGAACAATACGCTCCCATGTACAATCACAACTATAGGGGCAAACAATAACAATACAAGCAGTAATCGAACGAATACAAATAACAGC
 H  G  G  S  G  S  A  G  A  N  N  T  L  P  C  T  I  T  T  I  G  A  N  N  N  N  T  S  S  N  R  T  N  T  N  N  S
AGTATTATTAATAGCTCTAATAACAACACTGGTGCTGATCATAACGGTGATAACAATGGTCATTATATTAATAGTCTAACAGATAATCTCACTGCGCCCGCGCCGTCCGAA
 S  I  I  N  S  S  N  N  N  T  G  A  D  H  N  G  D  N  N  G  H  Y  I  N  S  L  T  D  N  L  T  A  P  A  P  S  E
TTAAGCGGCACCAACCGTAGAAAAATAATGGACATTGGCTTCGCGAGTAGCACTCATGTAAACATTGCTGCTCACTCTGGCCGTAAAGAAGCTGCACTGACGTCGCTCTCA
 L  S  G  T  N  R  R  K  I  M  D  I  G  F  A  S  S  T  H  V  N  I  A  A  H  S  G  R  K  E  A  A  L  T  S  L  S
TCGTCGCCCGCCACAGGCAATGCGGACAGTCAAAACGGTGACGCCGCAGGTAGTATTGATAATATAGGGGGTAGTAGCAACATCACGACCAACAGCAGTTTCAGTGGCAGC
 S  S  P  A  T  G  N  A  D  S  Q  N  G  D  A  A  G  S  I  D  N  I  G  G  S  S  N  I  T  T  N  S  S  F  S  G  S
GTTACAAACGCAACTGCACGCGCCAATAGCGTCAGTGAACCAACTATCCGTCCACGGGAGCCTCGCTACCGGCCCATGGTCGATACTGTTGTAGGCTCCATGGTTCTCGGG
 V  T  N  A  T  A  R  A  N  S  V  S  E  P  T  I  R  P  R  E  P  R  Y  R  P  M  V  D  T  V  V  G  S  M  V  L  G

GCCATTCTCGTTATTGGAGTGGCCGGCAACGCCATGGTCGTTGCTGTGGTCGCCCAGACGCGCTCGATGCGGACACCAACAAACTGTTACCTCGTCTCGTTATCAGTTGCC
 A  I  L  V  I  G  V  A  G  N  A  M  V  V  A  V  V  A  Q  T  R  S  M  R  T  P  T  N  C  Y  L  V  S  L  S  V  A

GATCTCATAGTACTTCTGTCAGCGATTCCGAACGAGATTCTCGCCCAGTACATCCTCGAGGATGAGTGGGTCTTTGGTCGGACGGGCTGCGCGCTATTCGTTTTCTTCCAG
 D  L  I  V  L  L  S  A  I  P  N  E  I  L  A  Q  Y  I  L  E  D  E  W  V  F  G  R  T  G  C  A  L  F  V  F  F  Q

TACCTGGGCATCAATGTCTCGTCGCTGTCAATTACTGCGTTCACCGTCGAACGCTATATAGCCATCTGTTTACCGATGAAGGCACAGACGATGTGCACGGTGAAGCGTGCT
 Y  L  G  I  N  V  S  S  L  S  I  T  A  F  T  V  E  R  Y  I  A  I  C  L  P  M  K  A  Q  T  M  C  T  V  K  R  A

AAGAAGATAATTCTTGCCGTGTGGTTTTTTGCTCTGTGTTATTGCTGTCCGTGGCTGTTCTTCCTCACAACGACGAAGCCGATTCACTATCGCGGATACGAGCACCGTTAC
 K  K  I  I  L  A  V  W  F  F  A  L  C  Y  C  C  P  W  L  F  F  L  T  T  T  K  P  I  H  Y  R  G  Y  E  H  R  Y

ATAGAGACATGCACGTATTCCTTACATAGATCCTTCTATATGCGGTTCTTTTTTGCTGACATCATCCTATTCTACGTCATTCCACTCATTCTGTCCTGTGTCCTGTACGGA
 I  E  T  C  T  Y  S  L  H  R  S  F  Y  M  R  F  F  F  A  D  I  I  L  F  Y  V  I  P  L  I  L  S  C  V  L  Y  G

CGAATGGCTCGAGTCCTATTCCATACGGACATTGGTGCTATGGCCTGTAAGGCTAATGGAGAGGCTAAGGCAAACGCGAATAATCAGTCGCGTATTCAGGTGGTGAAGATG
 R  M  A  R  V  L  F  H  T  D  I  G  A  M  A  C  K  A  N  G  E  A  K  A  N  A  N  N  Q  S  R  I  Q  V  V  K  M

CTGATCGTGGTAGTTGTGCTCTTTGCAACATTGTGGCTACCGTACCGTGTACTGCTCGTCTACAATTCGTTCGCCAGGAAGCCCTACCTGGAACTGTGGTACCTGATGCTC
 L  I  V  V  V  V  L  F  A  T  L  W  L  P  Y  R  V  L  L  V  Y  N  S  F  A  R  K  P  Y  L  E  L  W  Y  L  M  L

TGCAAGACGCTCATCTTCGTCAATTCGGCGATCAATCCTGTACTCTACAACGCCATGTCAGCCAAGTTTCGGCGCGCTTTCCGACGTACCCTATCATGCGGCTGCGGGTGT
 C  K  T  L  I  F  V  N  S  A  I  N  P  V  L  Y  N  A  M  S  A  K  F  R  R  A  F  R  R  T  L  S  C  G  C  G  C

GGAGATGTACGCTTGGAGCAACGAGGTTTTTCAACAGTCATGACCGTCCGCCAGCAAACGAATTTGCATCGTCCTTCGCGAAAGGATCAATACTAAGGACGCCCTCAAATC
 G  D  V  R  L  E  Q  R  G  F  S  T  V  M  T  V  R  Q  Q  T  N  L  H  R  P  S  R  K  D  Q  Y  *
ACTAGGACCCGTTCCCTTGATGCAGACGGATCAACGAATCGAGTCAAGGGAAGGAGCTATAATTTGCTTTTCCATTGAGAAAAAAATATATAGCGTAAAAGGTTAAATATT
TTTTGGTTAGCGCGAAGATCTGATGAAGCGATGAAATTGCTTGGTCGTCTTCAATTTTTGATATATATTAAAAGTAATATGTAAAATGAAGGGATACCTGGCGTGCATGAG
TAAAGTAAGTCTGTCCAACACAGAAGCAGCTCTATCGGGCTTTTTTTCTATACATCCAAACACAAAGAAACACCTTGTAGGTAAGTCAAACAGTTGCTTTTATGCATTAAT
CATACACTGGCCGTATCATTTACAACTGACGCGGCTCTCTTTGCTACCAGCGAGCCGCTCCGTATATGCGTATAAACTATATCGCATGTTCCGAGCTCTGAAACTGACGTT
AGCAATGATTTTAGACTATTAAGCCAGATAATTAGAGTCCATGCTGGAGATCCACAATGTTCCATTCTATTAAAACAGGCAATTACCTAACATAATGTCCTAAGTTACTTT
TGAGCCGGAAGGAAATTCAAATTAAAAGCCACCTTTCCCTCCTTTAGCGTCCTTGTCAAGCTACTCTTGTCCTTTTACTAAATGCGTATGAATATGAATACCGGCCTATGT
TTCTAGCACCAACGCTACAGCGTTCTAATTGACTCTGGCTAAATACTACCGCCCACGTACCGGGAACTGGTTTATAACATGACATTGTTTTACAGACCAATTCGCAGATTC
GAGAATAGAAAGCCATGTAATATAGATAGGATTGTACCATAAGCGGATATTATAGGCAACAAAGGGCACAAGGCGAAGCGAGCATTTCATTGAAGGTCCAGGGCTGATGTT
GAACATTGATGCTGTGTATAACCGCCAAGAGACGAACGCCGAGTGTCGTGTCTGTCGCATTCTGTGGTCCTTGTCTTGGTTGGCCATGGTAATTTTCTTAGAGGGTCGACG
ATATAACCCTCCCGCATGTATAAACACGGTGTAAAGATAAAATCTCGTCAATATGATTTGATGTATTGACTAAAGAATAAAATGAAATCAGAATAATGCACCGAGCGTGGC
AGTCGGAAATGGAATATTGCAAGCAAGAAACTGCGTTATGAGTTGATTGTCTAGGATGGGTCCTTTTACTGTCACAAACGGAAGGATATATACACTAAATGACAGGCTATA
TATATAAATATATAAAAAGATATATATATGTGCATGGTAGTGATAATGCTATGTGTTATTTCGTGAAGTAATGTTGGAAAGAATGATATGGAAAGCTTCGATGGTA
```

V. jacobsoni

```
CAAAACCGTAACAATAGCAGCGAACAACGAGCAACACCGAAATTAAAGACATCAATAAATGTGTAACCTCGAACAAATTAACAAGCGACGATTGAGTGTGTTAACAACTTG
GTGGCGGACGCTGTCGACACCGTTACTATTTAACAGTTTTCATTATTCGGTTCCGATAAAGTACCGTTTCACAGAAATCATTGTTAGACGCTTTCTCGTTGTCCTCTCTGT
GTGAAAGCAATAGCAACGAGGAAAAAGAATAAAAGACCTAGGATTATAGACTACACAGTGGTCGAGCGCTGCTTCCCATGTGGAGCAGTTGTCTTGACCGAAAACGAGCGG
TGTTTTGTCAAAGACACTTAGCAGTTAACTCACGACCCATGAAAACTTTGCCGCTAGCCAAACGTTCCGCCTATCTGTCCAGTCGGTCATAGGCAGTCGCACCAACCATGA
                                                                                                    M
AGTGGCCCATCGTCGCATCTACCTGGCTCAGTGTTGCTGTTGCCAACGTGGTACCAGTTGTGGACCCATTGTGTTCGACGAATATTGCCAACAACAACAGGACCAGCATAC
K  W  P  I  V  A  S  T  W  L  S  V  A  V  A  N  V  V  P  V  V  D  P  L  C  S  T  N  I  A  N  N  N  R  T  S  I
AAATGGCAATTGATTGTGACGACAGCGACTTGTACGACGCTATCAATGGAGGTTCCGTAACGAATTCGAACACCTCTATCGGTTCCTGGTCCTCTTCAGCCTCATCGATGG
```

```
Q M A I D C D D S D L Y D A I N G G S V T N S N T S I G S W S S S A S S M
CCTCGTCGTCGGCGGCGTCGAATGGGGTGCTGACAGGCAGCCAGAGTGTCAATGCCACGAGCTCTGTTAGGGCGAATTCACTGACGAGGGAGCCGAGGTATCGTCCCTTGG
A S S S A A S N G V L T G S Q S V N A T S S V R A N S L T R E P R Y R P L
TAGACACCATCGTTGGGACGGCGGTTCTTGGCGCCATCCTTCTTATCGGAGTTGCAGGCAATATAATGGTTGTTGCAGTGGTCGCCCAAACGCGCTCAATGCGAACTCCGA
V D T I V G T A V L G A I L L I G V A G N I M V V A V V A Q T R S M R T P

CGAACTGCTATCTCGTCTCGTTGTCAGTAGCCGACCTCATGGTTCTGCTATCGGCAATCCCGAATGAGATCCTCGCCCAGTACATCCTTGAGGACGAATGGATCTTTGGCC
T N C Y L V S L S V A D L M V L L S A I P N E I L A Q Y I L E D E W I F G

GAGCTGGCTGTGCAATGTTTGTTTTTTTCCAGTACCTGGGCATTAACGCCTCGTCTTTATCAATTACGGCATTTACCGTCGAGCGATATATCGCCATCTGCCTGCCAATGA
R A G C A M F V F F Q Y L G I N A S S L S I T A F T V E R Y I A I C L P M

AGGCGCAGACCATGTGCACGGTGAAACGTGCTAAGAAGATTATCCTCGCTGTGTGGTTGTTTGCGCTGTGTTACTGTTGCCCATGGTTATTTTTTCTTACGACGACACGGC
K A Q T M C T V K R A K K I I L A V W L F A L C Y C C P W L F F I T T T R

CAATCTACTACCGGGGCCAGGAGCACCGTTACGTGGAGACGTGCACGTATTCACTGCACCGATCGTTCTATATGCGTTTTTTTTTTGCCGACATCATTCTATTCTATGTAA
P I Y Y R G Q E H R Y V E T C T Y S L H R S F Y M R F F F A D I I L F Y V

TTCCGCTCATTCTATCATGTGTCTTGTACGGGCGAATGGCCCAAGTGTTATTCCATACGGACATTGGTGCGATGACGACGGCAGCGAAGACTGCCAACGGGCAGACTACAG
I P L I L S C V L Y G R M A Q V L F H T D I G A M T T A A K T A N G Q T T

CTAAAGTCAATGCTAATATGAATAACCAGTCCAGAATTCAGGTGGTGAAAATGCTGATCGTCGTCGTTGTACTTTTCGCGACGTTGTGGCTACCATATCGTGTGCTGCTCG
A K V N A N M N N Q S R I Q V V K M L I V V V V L F A T L W L P Y R V L L

TCTATAATTCGTTTGCCAGCAAGCCTTACCTCGAGCTGTGGTACCTGATGCTCTGCAAGACGCTCATCTTCGTCAATTCGGCCATCAATCCTGTGCTCTACAACGCCATGT
V Y N S F A S K P Y L E L W Y L M L C K T L I F V N S A I N P V L Y N A M

CGGTCAAGTTCCGGCGGGCCTTTCGGCGGACCTTATCCTGCGATTCTAAAAATCGTTTCGAAAATCGCACGTTTACTACAACAATGACAACGGTCAGGCACACGATAAAGC
S V K F R R A F R R T L S C D S K N R F E N R T F T T T M T T V R H T I K
CGAAGGATCAATATTGACGGGACGCGAACTCCTAAATTACTTTTGCCTAGAAGCTCGCAGAAATGAGGCAAACAACCGATTTCGAGAAGTACCTCGTCGTAAATGGCAACA
P K D Q Y *
CCTCAACGCCGAACGACACAGAAAATTCGTCGAAAATTGCGCTTGCCGATAAACCTAACCAAGAGAATAAGAATGAATGCAACCTGCAAGCTGCGACTCACCATTGTTAAA
CGACTTCACATCCCCCTTAGGCAGCCCTGTTTCGAAGATATTTTGTTGCAGAAAATCGCGATGATAACTCAGTTTTACCCTGCACAACGTTGATAGTATTGCTGACCCAAT
GCTAAAACAGCTAAATCAATGAGTTCACAAAAGTCATGGCCGCGATTTTTCAACGTAGGGACCTGTAACTGGTCGTCATTGAATGAGTAAATTGAAAGAAAAACATTAAAT
TGGAATTTTCCTGACTGAGTATAATTATAGTAGTATCGATAAAGTGAATTAATGTACACTAAATATGAATCTATCGTATACATTTTGGTGGAATGTAGCACTTAAATGTGC
TTTGAGCACAAAAAAAAACAAGCAATAGATATCAATGCCTGATATCCTAAAATTATAGTTTGTACCGTTGATGCAATAGACCATAACTGCATTGGCCATTGGGTCGTTGTG
CTTCTGCAGACTGCAGAATCGTCCAATGTGTGCAGATTCTCTGATTCGCTTCACTGACGTATTGCGTTTTGCCAGCCGCCAAGGTGCACTCAAACAAAGTCTACCATAAGT
CGACTTCTGCGTATTTATAATCATATACCCTCACATCGTGTACTTTAACATCCTGTTTGTGTCGTACCAACAAACAAATCAACTTCCTACACGTTTCGCGTAGCTTCCGCC
AATAGCCCAGCCCGCGCAATGAAATCAGTCAATAAAACAAAACCCGAATGCGCGAGATGCACCCAAGACGAATCGAGAGTCTTTTATCTGCAGTTAGAAAGGCAAACGAAG
ACTCTTTTGTTTGCTTCCGACGTGAGAATTCCCGTACCACGATATTATTCGTAAGATGAGTGGATCTGGGAGGAGTCGGGGAAAATGTTAACCAGCGGGAATAATTGAACG
AGACCGAAGTAAAATCGAATTGAAGTTAGACCGAGTTTATTTTGCTTCCAGTAATCAGTACTACTATTGCTTATGCTAATAGAATACTAAAGGAAATAGATTCAGCCTCTC
TTAGTTGTTTCGTGAGAACGAATTTTACGTGCAGACGCATAGACCAACACACACACACACGTACACGCCCATGCCGGCATGAAAGGATAATAAACAGCATAGAATTATG
ATATCTGTATGGGATGCAACTGCGGGCGTCGTCGGAAAAAAACAAAAGGTAAGAACTACAATCACATTTTGTAGATTAAGAGCAACACGCCATGTTTTCATGGTATCAATT
CAGACAAGTATAGAGAGTATAAAAAAAAAATGTGATCTAATTAACATTTTTCACTTGGCCAGGCGGCGAAATGGACACTCAAGGCATGTTAAAGTTACTAACTAAGACAATT
TCTTTCAAAAAGCTAGACTGTTTACGAAGTCCAGAACTCTTCACAGCGGAAGCAGGTGGGTGGTGTAGCAGCTTGAGGTAGTTCATCTATGCATACACATAAACCACTAGG
ACCGTATAACGGACCATTGTAGCGGCAGTACATGCAAATGTAAGATGAATGCAGCCTTAACACTGTCGATTGACACAGACATTTAATTTTAATCTGTATCTTTAAAAATAT
TTTCTAGTACAGTTTAGAGTAAAAAAAAATAGTCCTCTTCCTATGGCGGGATCAGATTATCATAAGTAGAGATTCCGAGCAAGGAACATACGACGAATATGTTGAAAACATA
CAATGGATCTCAGTAAATCCGAAGATGACTTTTGCCGTGACAAATGGCTCTATCTATTGTGCATGAGAGCAGAAAACCCGTTGAGGCATTTCCTATTAACTCCGGGAAAGA
ATAGGCCGGCTTTGAGCAAGGCACCAACAATCTGTAAACGACACCAGTCTGTCACACAGTTTTGTAAAACGGCTAAAAAAAAAAATACGGGGTACATCTACGAGACAGAGT
AACGATCAGACTCTCTTCTAGAGTTACCCACAACATGTAAAAAAAACTGATTAAAGAAGAGCAGTTTTTCATAAGATATTTTACCTTATTGTACAGTATGATGTTATAAAC
GAGGGTTGATGAATATATAAATGACATTGAAAAAAAAACAGTCGAACCTGGTGTATATATTTATTCTTCTCACTGAAAAACGCCTCCATATGTAAATGTAATAGAATAATC
CGAGCAGGAGGTAGAGGTGGACAGGGTAAGAAAGGGCAAGAAACGAGAGTGTCCGATGAAAGAGAAATGGTCTCTGTGCTGTGGTGTTAATAAAAGCGTCAAAATATTTGT
CGGGTTTGTGAAAATATATATTAAAATGATGACCATGATAACGAGGATAAATAAAATGACGAATTACCTTGAA
```

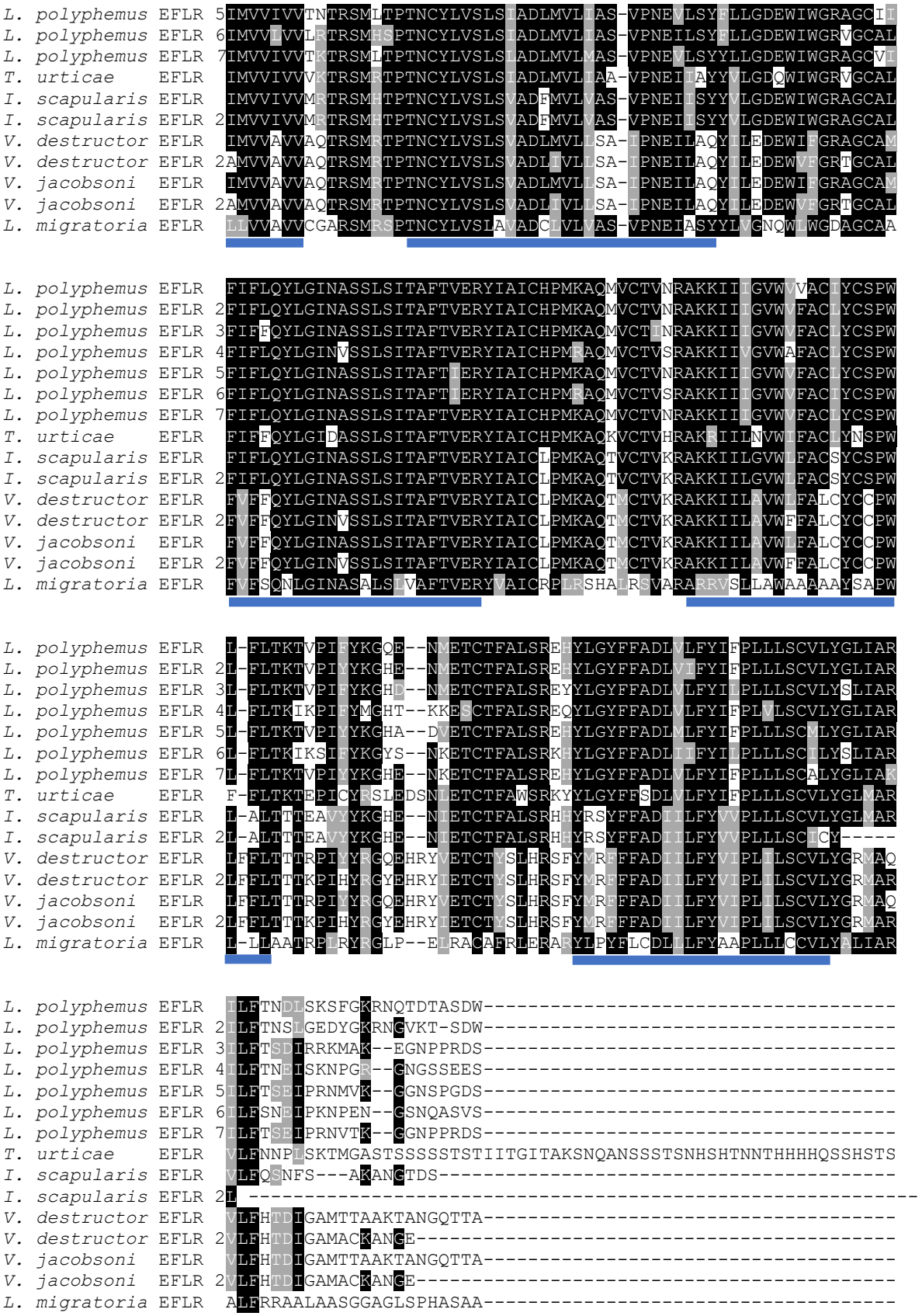
EFLR2:

```
ATTCGATCGTTATTGGAGTTTTAAAGTACTGCGAAGAAGTTCTTAAGAGTTCACCTGTTTCAACGGGCCCAGCCATACGCGTACTCTTCTATCCATTAGCTAACACAGTGC
CGTGTCGCCGCTTTTGTGGTTTCGTGCACGCATGTGTATGTGACATGGCGATTACCATTGAACTCGATCATTTGTTTGGATGTTTGTTACTCGTGCCCCTGGACGTAACCT
AGCAGTCTAACGAGTCGAGAAAGATCTTGTACTGGTTTTTCAGAAAACAGAAACGAAACCATGTCATATCCAGGCTCCGAAACAATTTTAGATGACTACGAAATAATAGCT
ACAACCATGTCGTCTTCTTCTTTTGATTGTTGCTACATGCGCCATGTCTACAATGTGATAGTGTGTCAGCCTGTGCGTTTGTATAATTCACCACTACCCAGCCACGTTTAA
CTTAGGTCGGCTGAAATTTCAAAGAAAACCGAATCTATCGAAGATCTCGTTTCCGAGTGCTTTCACATTCGATTATACTCGTTAGTTCTGTGCTATTTTCTGTTGTGCCCA
GAGCCGTCGCTATCGCATGATGTGCTGCTTCGGCGAAAAAGTCGCTCCGGCGTTCTGTCACGTAAGACGACGGATGCTGCTACTAGTAACATTATTTCTTGTCATTGATGC
                M M C C F G E K V A P A F C H V R R R M L L L V T L F L V I D A
GCGAACGGTTAAGTCGACATGCGTCGGCGGAAAAGCACATGGCGGCAGCGGAAGCGCAGGCGCGAACAATACGCTCCCATGTACAATCACAACTATAGGGGCAAACAATAA
  R T V K S T C V G G K A H G G S G S A G A N N T L P C T I T T I G A N N N
CAATACAAGCAGTAATCGAACGAATACAAATAACAGCAGTATTATTAATAGCTCTAATAACAACACCGGTGCTGATCATAACGGTGATAACAATGGTCATTATATTAATAG
  N T S S N R T N T N N S S I I N S S N N N T G A D H N G D N N G H Y I N S
TCTAACAGATAATCTTACTGCGCCCGCGCCGTCCGAATTAAGCGGCACCAACCGTAGGAAAATAATGGACATTGGCTTCGCGAGTAGCACTCATGTAAACATTGCTGCTCA
  L T D N L T A P A P S E L S G T N R R K I M D I G F A S S T H V N I A A H
CTCTGGCCGTAAAGAAGCTGCACTGACGTCGCTCTCATCGTCGCCCGCCACAGGCAATGCGGACAGTCAAAACGGTGACGCCGCAGGTAGTATTGATAATATAGGGGGTAG
```

```
  S  G  R  K  E  A  A  L  T  S  L  S  S  S  P  A  T  G  N  A  D  S  Q  N  G  D  A  A  G  S  I  D  N  I  G  G  S
TAGCAACATCACGACCAACAGCAGTTTCAGTGGCAGCGTTACAAACGCAACTGCACGCGCCAATAGCGTCAGTGAACCAACTATCCGTCCACGGGAGCCTCGCTACCGGCC
  S  N  I  T  T  N  S  S  F  S  G  S  V  T  N  A  T  A  R  A  N  S  V  S  E  P  T  I  R  P  R  E  P  R  Y  R  P
CATGGTCGATACTGTTGTAGGCTCCATGGTTCTCGGGGCCATTCTCGTTATTGGAGTGGCCGGCAACGCCATGGTCGTTGCTGTGGTCGCCCAGACGCGCTCGATGCGGAC
  M  V  D  T  V  V  G  S  M  V  L  G  A  I  L  V  I  G  V  A  G  N  A  M  V  V  A  V  V  A  Q  T  R  S  M  R  T

ACCAACAAACTGTTACCTCGTCTCGTTATCAGTTGCCGATCTCATAGTACTTCTGTCAGCGATTCCGAACGAGATTCTCGCCCAGTACATCCTCGAGGATGAGTGGGTCTT
  P  T  N  C  Y  L  V  S  L  S  V  A  D  L  I  V  L  L  S  A  I  P  N  E  I  L  A  Q  Y  I  L  E  D  E  W  V  F

TGGTCGGACGGGCTGCGCGCTATTCGTTTTCTTCCAGTACCTGGGCATCAATGTCTCGTCGCTGTCAATTACTGCGTTCACCGTCGAACGCTATATAGCCATCTGTTTACC
  G  R  T  G  C  A  L  F  V  F  F  Q  Y  L  G  I  N  V  S  S  L  S  I  T  A  F  T  V  E  R  Y  I  A  I  C  L  P

GATGAAGGCACAGACGATGTGCACGGTGAAGCGTGCTAAGAAGATAATTCTTGCCGTGTGGTTTTTTGCTCTGTGTTATTGCTGTCCGTGGCTGTTCTTCCTCACAACGAC
  M  K  A  Q  T  M  C  T  V  K  R  A  K  K  I  I  L  A  V  W  F  F  A  L  C  Y  C  C  P  W  L  F  F  L  T  T  T

GAAGCCGATTCACTATCGCGGATACGAGCACCGTTACATAGAGACATGCACGTATTCCTTACATAGATCCTTCTATATGCGGTTCTTTTTTGCTGACATCATCCTATTCTA
  K  P  I  H  Y  R  G  Y  E  H  R  Y  I  E  T  C  T  Y  S  L  H  R  S  F  Y  M  R  F  F  F  A  D  I  I  L  F  Y

CGTCATTCCACTCATTCTGTCCTGTGTCCTGTACGGACGAATGGCTCGAGTCCTATTCCATACGGACATTGGTGCTATGGCCTGTAAGGCTAATGGAGAGGCTAAGGCAAA
  V  I  P  L  I  L  S  C  V  L  Y  G  R  M  A  R  V  L  F  H  T  D  I  G  A  M  A  C  K  A  N  G  E  A  K  A  N

CGCGAATAATCAGTCGCGTATTCAGGTGGTGAAGATGCTGATCGTGGTAGTTGTGCTCTTTGCAACATTGTGGCTACCGTACCGTGTACTGCTCGTCTACAATTCGTTCGC
  A  N  N  Q  S  R  I  Q  V  V  K  M  L  I  V  V  V  V  L  F  A  T  L  W  L  P  Y  R  V  L  L  V  Y  N  S  F  A

CAGGAAGCCCTACCTGGAACTGTGGTACCTGATGCTCTGCAAGACGCTCATCTTCGTCAATTCGGCGATCAATCCTGTACTCTACAACGCCATGTCAGCCAAGTTTCGGCG
  R  K  P  Y  L  E  L  W  Y  L  M  L  C  K  T  L  I  F  V  N  S  A  I  N  P  V  L  Y  N  A  M  S  A  K  F  R  R

CGCTTTCCGACGTACCCTATCATGCGGCTGCGGGTGTGGAGATGTACGCTTGGAGCAACGAGGTTTTTCAACAGTCATGACCGTCCGCCAGCAAACGAATTTGCATCGTCC
  A  F  R  R  T  L  S  C  G  C  G  C  G  D  V  R  L  E  Q  R  G  F  S  T  V  M  T  V  R  Q  Q  T  N  L  H  R  P
TTCGCGAAAGGATCAATACTAAGGACGCCCTCAAATCACTAGGACCCGTTCCCTTGATGCAGACGGATCAACGAATCGAGTCAAGGGAAGGAGCTATAATTTGCTTTTCCA
  S  R  K  D  Q  Y  *
TTGAGAAAAAAATATATAGCGTAAAAGGTTAAATATTTTTTGGTTAGCGCGAAGATCTGATGAAGCGATGAAATTGCTTGGTCGTCTTCGATTTTTGATATATATTAAAAG
TAATATGTAAAATGAAAGGATACCTGGCGTGCATGAGTAAAGTAAGTCTGTCCAACACAGAAGCAGCTCTATCGGGCTTTTTTTCTATACATCCAAACACAAAGAAACACC
TTGTAGGTAAGTCAAACAGTTGCTTTTATGCATTAATCATACACTGGCCGTATCATTTACAACTGACGCGGCTCTCTTTGCTACCAGCGAGCCGTTCCGTATATGCGTATA
AACTATATCGCATGTTCCGAGCTCTGAAACTGACGTTAGCAATGATTTTAGACTATTAAGCCAGATAGTTAGAGTCCATGCTGGAGATCCACAATGTTCCATTCTATTAAA
ACAGGCAATTACCTAACATAATGTCCTAAGTTACTTTTGAGCCGGAAGGAAATTCAAATTAAAAGCCACCTTTCCCTCCTTTAGCGTCCTTGTCAAGCTACTCTTGTCCTT
TTACTAAATGCGTATGAATATGAATACCGGCCTATGTTTCTAGCACCAACGCTACAGCGTTCTAATTGACTCTGGCTAACTACTACCGCCCACGTACCGGGAACTGGTTTA
TAACATGACATTGTTTTACAGACCAATTCGCAGATTCGAGAATAGAAAGCCATGTAATATAGATAGGATTGTACCATAAGCGGATATTATAGGCAACAAAGGGCACAAGGC
GAAGCGAGCATTTCATTGAAGGTCCAGGGCTGATGTTGAACATTGATGCTGTGTATAACCGCCAAGAGACGAACGCCGAGTGTCGTGTCTGTCGCATTCTGTGGTCCTTGT
CTTGGTTGGCCATGGTAATTTTCTTAGAGGGTCGACGATATAACCCTCCCGCATGTATAAACACGGTGTAAAGATAAAATCTCGTCAATATGATTTGATGTATTGACTAAA
GAATAAAATGAAATCAGAATAATGCACCGAGCGTGGCAGTCGGAAATGGAATATTGCAAGCAAGAAACTGCGTTATGAGTTGATTGTCTAGGATGGGTCCTTTTACTGTCA
CAAACGGAAGGATATATACACTAAATGACAGGCTATATATATAAATATATAAAAAGATATATATATGTGCATGGTAGTGATAATGCTATGTGTTATTTCGTGAAGTAATGT
TGGAAAGAATGATATGGAAAGCTTCGATGGTAAATAAAAAACGGTTACTTAAATAG
```

B) Multiple sequence alignment:

```
L. polyphemus EFLR  ------------------------NLTDY-AAPTYYSINYRIVGTIFQGIIFVVGALGN
L. polyphemus EFLR 2------------------------KIIAY-QDPTYYSINYRIIGTIFQGVIFLVGVLGN
L. polyphemus EFLR 3------------------------NTTIALPEPSYYSIYYRIIGTTFLGLIFLIGILGN
L. polyphemus EFLR 4----------------------FLASTTY-RDPSYYSRNYRIIGTFFQGIIFSVGVLGN
L. polyphemus EFLR 5------------------------NTTML-PEPSYYPIYYRVIGTTFQGLIFLIGVLGN
L. polyphemus EFLR 6--------------------------PDPSFYSVYYRIIGTLFQGFIFMTGALGN
L. polyphemus EFLR 7------------------------NMTMMLPEPSYYPAYYRIVGTTFQGLIFLIGVLGN
T. urticae    EFLR  PATGTASPAVVAAAAGFSPETYCNNLNTTFCKDPSYYAISYRIIGTIFQGFILIIGVLGN
I. scapularis EFLR  ----------------------------Y-QDPTYYSVNYRIIGTIFQGIIFGVGVLGN
I. scapularis EFLR 2----------------------------Y-QDPTYYSVNYRIIGTIFQGIIFGVGVLGN
V. destructor EFLR  --ASSSAASNGVLTGSQSVNATSSVRANSLTREPRYRPLVDTIVGTAVLGAILLIGVAGN
V. destructor EFLR 2SSNITTNSSFSGSVTNATARANSVSEPTIRPREPRYRPMVDTVVGSMVLGAILVIGVAGN
V. jacobsoni  EFLR  --ASSSAASNGVLTGSQSVNATSSVRANSLTREPRYRPIVDTIVGTAVLGAILLIGVAGN
V. jacobsoni  EFLR 2SSNITTNSSFSGSVTNATARANSVSEPTIRPREPRYRPMVDTVVGSMVLGAILVIGVAGN
L. migratoria EFLR  --------------------------------AEPAYYSARYRLVGTLCQGVVLAVGLAGN

L. polyphemus EFLR  IMVVIVVTKTRSMRTPTNCYLVSLSLADLMVLVAS-VPNEVLSYYLLGDEWIWGRAGCVI
L. polyphemus EFLR 2IMVVIVVTKTRSMLTPTNCYLVSLSLADLLVLMAS-VPNEILSYYLLGDEWIWGRAGCII
L. polyphemus EFLR 3IMVVVVVTKSRSMLTPTNCYLVNLSLADLMVLVAS-VPNEVLSYYLLGDEWIWGRVGCAL
L. polyphemus EFLR 4IMVVIVVRKTRSMRTPTNCYLVSLSIADLMVLTAS-VPNEILSYFLLGDEWIWGRVGCAL
```



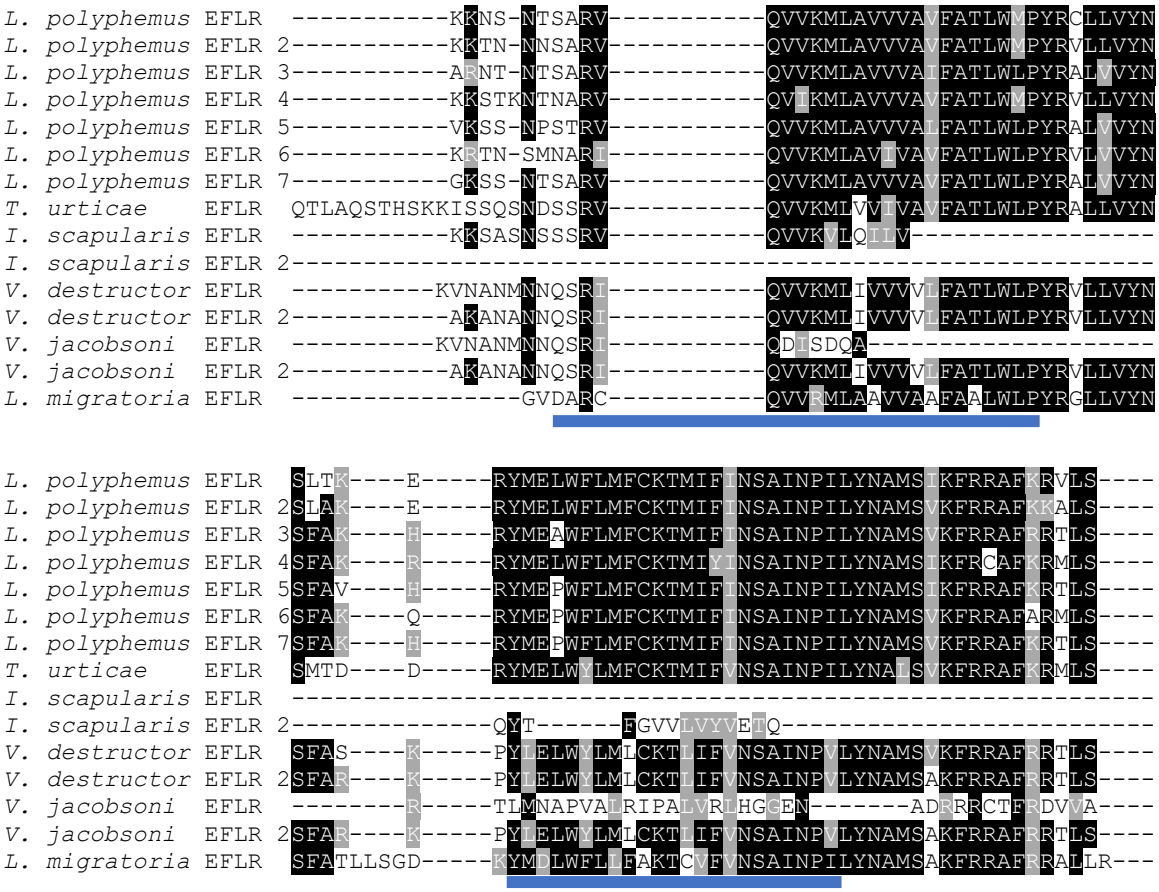
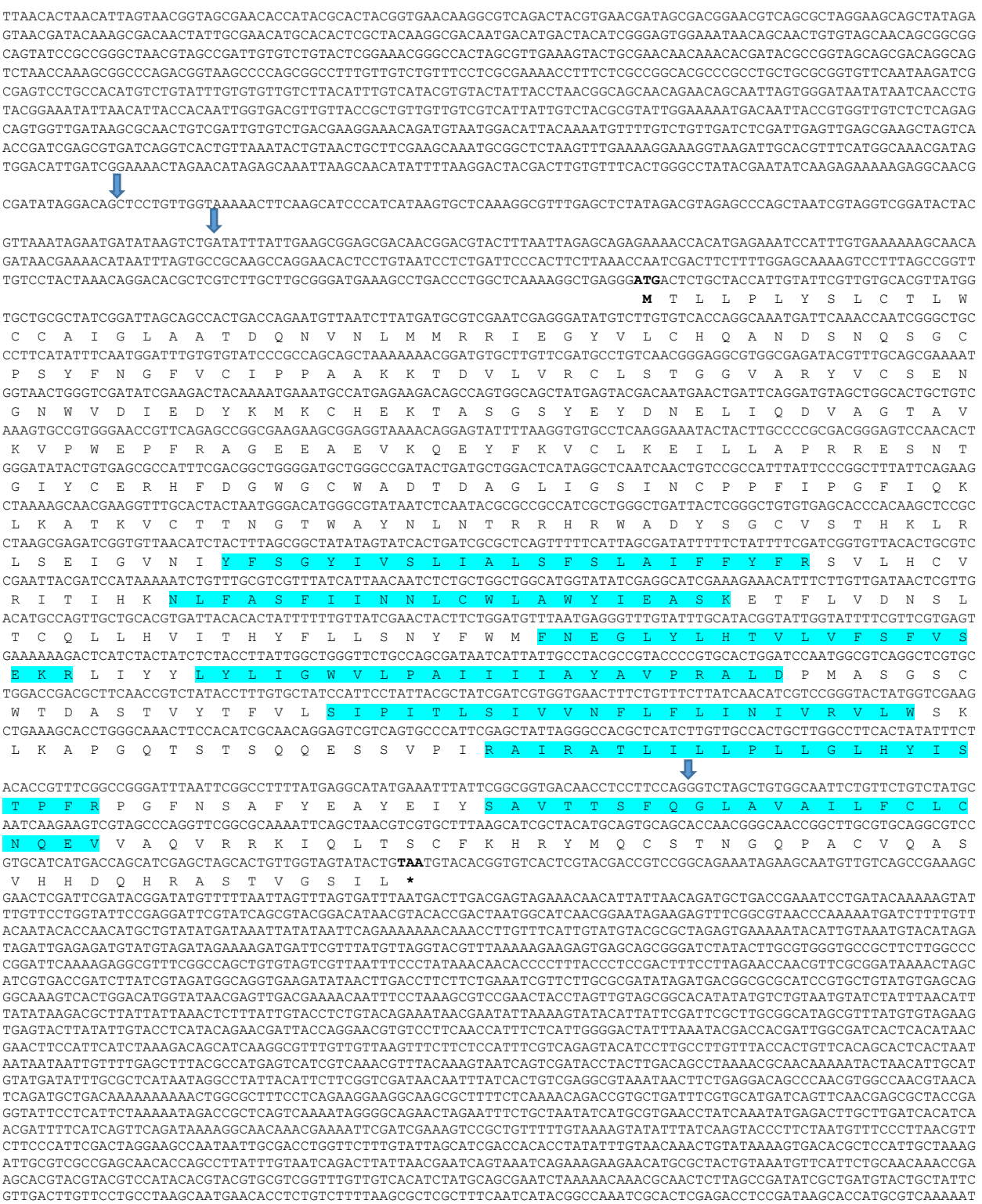


Figure 3.21 Gene structure for EFLamide receptor sequence in *V. destructor* and *V. jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of EFLamide receptor from various orthologous species across Phylum Arthropoda.

Diuretic hormone 31 type receptor:

A) Gene structure

V. destructor



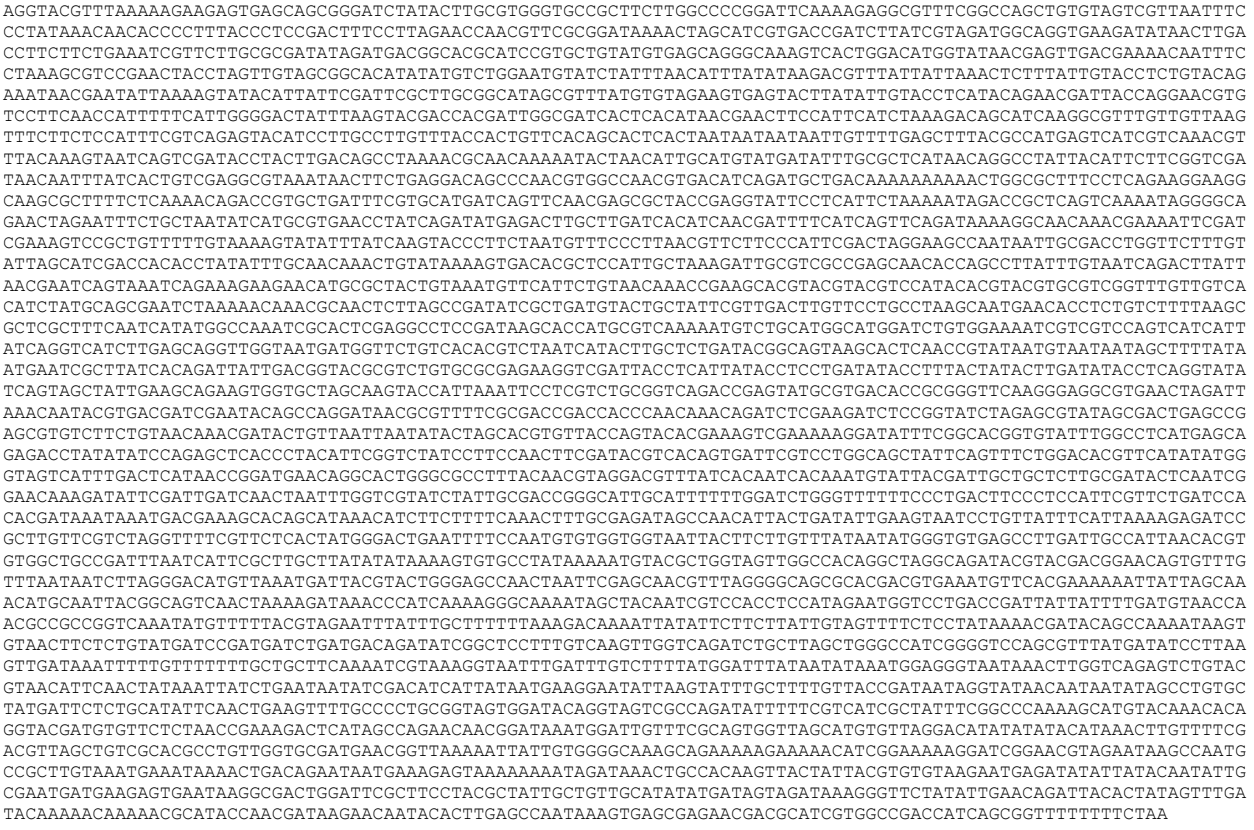
GTCTGCATGGCATGGATCTGTGAAAAATCGTCGTCCAGTCATCATTATCAGGTCATCTTGAGCAGGTTGGTAATGATGGTTCTGTCACACGTCTAATCATACTTGCTCTGA
TACGGCAGTAAGCACTCAACCGTATAATGTAATAATAGCTTTTATAATGAATCGCTTATCACAGATTATTGACGGTACGCGTCTGTGCGCGAGAAGGTCGATTACCTCATT
ATACCTCCTGATATACCTTTACTATACTTGATATACCTCAGGTATATCAGTAGCTATTGAAGCAGAAGTGGTGCTAGCAAGTACCATTAAATTCCTCGTCTGCGGTCAGAC
CGAGTATGCGTGACACCGCGGGTTCAAGGGAGGCGTGAACTAGATTAAACAATACGTGACGATCGAATACAGCCAGGATAACGCGTTTTCGCGACCGACCACCCAACAAAC
AGATCTCGAAGATCTCCGGTATCTAGAGCGTATAGCGACTGAGCCGAGCGTGTCTTCTGTAACAAACGATACTGTTAATTAATATACTAGCACGTGTTACCAGTACACGAA
AGTCGAAAAAGGATATTTCGGCACGGTGTATTTGGCCTCATGAGCAGAGACCTATATATCCAGAGCTCACCCTACATTCGGTCTATCCTTCCAACTTCGATACGTCACAGT
GATTCGTCCTGGCAGCTATTCAGTTTCTGGACACGTTCATATATGGGTAGTCATTTGACTCATAACCGGATGAACAGGCACTGGGCGCCTTTACAACGTAGGACGTTTATC
ACAATCACAAATGTATTACGATTGCTGCTCTTGCGATACTCAATCGGAACAAAGATATTCGATTGATCAACTAATTTGGTCGTATCTATTTCGATCGGGCATTGCATTTTT
TGGATCTGGGTTTTTTTCCTGACTTCCCTCCATTCGTTCTGATCCACACGATAAATAAATGACGAAAGCACAGCATAAACATCTTCTTTTCAAACTTTGCGAGATAGCCAA
CATTACTGATATTGAAGTAATCCTGTTATTTCATTAAAAGAGATCCGCTTGTTCGTCTAGGTTTTCGTTCTCACTATGGGACTGAATTTTCCAATGTGTGGTGGTAATTAC
TTCTTGCTTATAATATGGGTGTGAGCCTTGATTGCCATTAACACGTGTGGCTGCCGATTTAATCATTCGCTTGCTTATATATAAAAGTGTGCCTATAAAAATGTACGCTAG
TAGTTGGCCACAGGCTAGGCAGATACGTACGACGGAACAGTATTTGTTTAATAATCTTAGGGACTTGTTAAATGATTACGTACTGGGAGCCAACTAATTCGAGCAACGTTT
AGCGGCAGCGCACGACGTGAAATGTTCACGAAAAAATTATTAGCAAACATGCAATTACGGCAGTCAACTAAAAGATAAACCCATCAAAAGGGCAAAATAGCTACAATCGTC
CACCTCCATAGAATGGTCCTGACCGATTATTATTTTGATGTAACCAACGCCGCCGGTCAAATATGTTTTTACGTAGAATTTATTTGCTTTTTTAAAGACAAAATTATATTC
TTCTTATTGTAGTTTTCTCCTATAAAACGATACAGCCAAAATAAGTGTAACTTCTCTGTATGATCCGATGATCTGATGACAGATATCGGCTCCTTTGTCAAGTTGGTCAGA
TCTGCTTAGCTGGGCCATCGGGGTCCAGCGTTTATGATATCCTTAAGTTGATAAATTTTTGTTTTTTTGCTGCTTCAAAATCGTAAAGGTAATTTGATTTGTCTTTTATGG
ATTTATAATATAAATGGAGGGTAATAAACTTGGTCAGAGTCTGTACGTAACATTCAACTATAAATTATCTGAATAATATCGACATTATTATGAAGGAATATTAAGTAT
TTGCTTTTGTTACCGATAATAGGTATAACAATAATATAGCCTGTGCTATGATTCTCTGCATATTCAACTGAAGTTTTGCCCCTGCGGTAGTGGATACAGGTAGTCGCCAGA
TATTTTTCGTCATCGCTATTTCGGCCCAAAAGCATGTACAAACACAGGTACGATGTGTTCTCTAACCGAAAGACTCATAGCCAGAACAACGGATAAATGGATTGTTTCGCA
GTGGTTAGCATGTGTTAGGACATATATATACATAAACTTGTTTTCGACGTTAGCTGTCGCACGCCTGTTGGTGCGATGAACGGTTAAAAATTATTGTGGGGCAAAGCAGAA
AAAGAAAAACATCGGAAAAAGGATCGGAACGTAGAATAAGCCAATGCCGCTTGTAAATGAAATAAAACTGACAGAATAATGAAAGAGTAAAAAAAAATAGATAAACTGCCA
CAAGTTACTATTACGTGTGTAAGAATGAGATATATTATACAATATTGCGAATGATGAAGAGTGAATAAGGCGACTGGATTCGCTTCCTACGCTATTGCTGTTGCATATATG
ATAGTAGATAAAGGGTTCTATATTGAACAGATTACACTATAGTTTGATACAAAAACAAAAACGCATACCAACGATAAGAACAATACACTTGAGCCAATAAAGTGAGCGAGA
ACGACGCATCGTGGCCGACCAT

V. jacobsoni

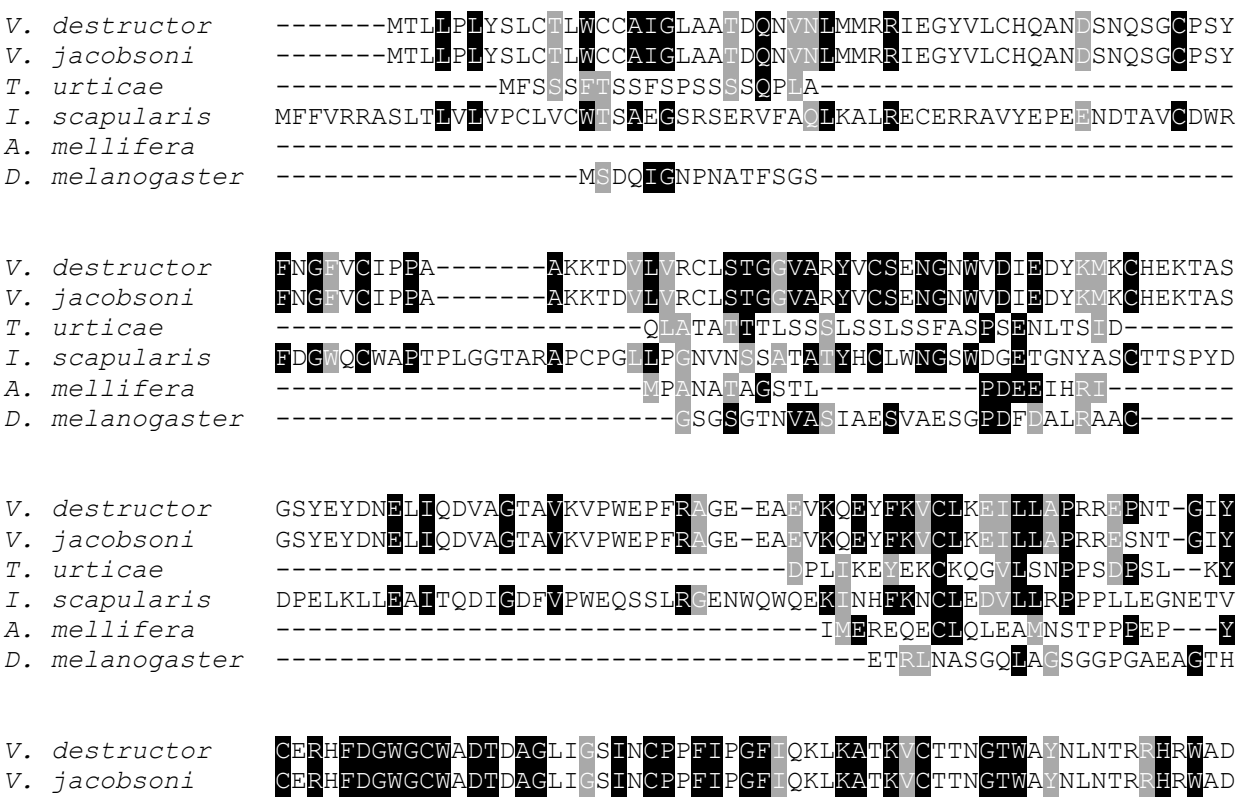
TCTTCATACGTCGACGAATTAACACTAACATTAGTAACGGTAGCGAACACCATACGCACTACGGTGAACAAGGCGTCAGACTACGTGAACGATAGCGACGGAACGTCAGCG
CTAGGAAGCAGCTATAGAGTAACGATACAAAGCGACAACTATTGCGAACATGCACACTCGCTACAAGGCGACAATGACATGACTACATCGGGAGTGGAAATAACAGCAACT
GTGTAGCAACAGCGGCGGCAGTATCCGCCGGGCTAACGTAGCCGATTGTGTCTGTACTCGGAAACGGGCCACTAGCGTTGAAAGTACTGCGAACAACAAACACAATACGCC
GGTAGCAGCGACAGGCAGTCTAACCAAAGCGGCCCAGACGGTAAGCCCCAGCGGCCTTTGTTGTCTGTTTCCTCGCGAAAACCTTTCTCGCCGGCACGCCCGCCTGCTGCG
CGGTGTTCAATAAGATCGCGAGTCCTGCCACATGTCTGTATTTGTGTGTTGTCTTACATTTGTCATACGTGTACTATTACCTAACGGCAGCAACAGAACAGCAATTAGTGG
GATAATATAATCAACCTGTACGGAAATATTAACATTACCACGATTGGTGACGTTGTTACCGCTGTTGTTGTCGTCATTATTGTCTACGCGTATTGGAAAAATGACAATTAC
CGTGGTTGTCTCTCAGAGCAGTGGTTGATAAGCGCAACTGTCGATTGTGTCTGACGAAGGAAACAGATGTAATGGACATTACAAAATGTTTTGTCTGTTAATCTCGATTGA

GTTGAGCGAAACTAGTCAACCGATCGAGCGTGATCAGGTCACTGTTAAATACTGTAACTGCTTCGAAGCAAATGCGGCTCTAAGTTTGAAAAGGAAAGATATTTATTGAAG
CGGAGCGACAACGGACGTACTTTAATTAGAGCAGAGAAAACCACATGAGAAATCCATTTGTGAAAAAAGCAACAGATAACGAAAACATAATTTAGTGCCGCAAGCCAGGAA
CACTCCTGTAATCCTCTGATTCCCACTTCTTAAACCAATCGACTTCTTTTGGAGCAAAAGTCCTTTAGCCGGTTTGTCCTACTAAACAGGACACGCTCGTCTTGCTTGCGG
GATGAAAGCCTGACCCTGGCTCAAAAGGCTGAGGG**ATG**ACTCTGCTACCATTGTATTCGTTGTGCACGTTATGGTGCTGCGCTATCGGATTAGCAGCCACTGACCAGAATG
M T L L P L Y S L C T L W C C A I G L A A T D Q N
TTAATCTTATGATGCGTCGAATCGAGGGATATGTCTTGTGTCACCAGGCAAATGATTCAAACCAATCGGGCTGCCCTTCATATTTCAATGGATTTGTGTGTATCCCGCCAG
V N L M M R R I E G Y V L C H Q A N D S N Q S G C P S Y F N G F V C I P P
CAGCTAAAAAAACGGATGTGCTTGTTCGATGCCTGTCAACGGGAGGCGTGGCGAGATACGTTTGCAGCGAAAATGGTAACTGGGTCGATATCGAAGACTACAAAATGAAAT
A A K K T D V L V R C L S T G G V A R Y V C S E N G N W V D I E D Y K M K
CGGAGGTAAAACGCCATGAGAAGACAGCCAGTGGCAGCTATGAGTACGACAATGAACTGATTCAGGATGTAGCTGGCACTGCTGTCAAAGTGCCGTGGGAACCGTTCAGAG
C H E K T A S G S Y E Y D N E L I Q D V A G T A V K V P W E P F R A G E E
CCGGCGAAGAAGAGGAGTATTTTAAGGTGTGCCTCAAGGAAATACTACTTGCCCCGCGACGGGAGCCCAACACTGGGATATACTGTGAGCGCCATTTCGACGGCTGGGGAT
A E V K Q E Y F K V C L K E I L L A P R R E P N T G I Y C E R H F D G W G
GCTGGGCCGATACTGATGCTGGACTCATAGGCTCAATCAACTGTCCGCCATTTATTCCCGGCTTTATTCAGAAGCTAAAAGCAACGAAGGTTTGCACTACTAATGGGACAT
C W A D T D A G L I G S I N C P P F I P G F I Q K L K A T K V C T T N G T
GGGCGTATAATCTCAATACGCGCCGCCATCGCTGGGCTGATTACTCGGGCTGTGTGAGCACCCACAAGCTCCGCCTAAGCGAGATCGGTGTTAACATCTACTTTAGCGGCT
W A Y N L N T R R H R W A D Y S G C V S T H K L R L S E I G V N I Y F S G
ATATAGTATCACTGATCGCGCTCAGTTTTTCATTAGCGATATTTTTCTATTTTCGATCGGTGTTACACTGCGTCCGAATTACGATCCATAAAAATCTGTTTGCGTCGTTTA
Y I V S L I A L S F S L A I F F Y F R S V L H C V R I T I H K N L F A S F
TCATTAACAATCTCTGCTGGCTGGCATGGTATATCGAGGCATCGAAAGAAACATTTCTTGTTGATAACTCGTTGACATGCCAGTTGCTGCACGTGATTACACACTATTTTT
I I N N L C W L A W Y I E A S K E T F L V D N S L T C Q L L H V I T H Y F
TGTTATCGAACTACTTCTGGATGTTTAATGAGGGTTTGTATTTGCATACGGTATTGGTATTTTCGTTCGTGAGTGAAAAAAGACTCATCTACTATCTCTACCTTATTGGCT
L L S N Y F W M F N E G L Y L H T V L V F S F V S E K R L I Y Y L Y L I G
GGGTTCTGCCAGCGATAATCATTATTGCCTACGCCGTACCCCGTGCACTGGATCCAATGGCGTCAGGCTCGTGCTGGACCGACGCTTCAACCGTCTATACCTTTGTGCTAT
W V L P A I I I I A Y A V P R A L D P M A S G S C W T D A S T V Y T F V L
CCATTCCTATTACGCTATCGATCGTGGTGAACTTTCTGTTTCTTATCAACATCGTCCGGGTACTATGGTCGAAGCTGAAAGCACCTGGGCAAACTTCCACATCGCAACAGG
S I P I T L S I V V N F L F L I N I V R V L W S K L K A P G Q T S T S Q Q
AGTCGTCAGTGCCCATTCGAGCTATTAGGGCCACGCTCATCTTGTTGCCACTGCTTGGCCTTCACTATATTTCTACACCGTTTCGGCCGGGATTTAATTCGGCCTTTTATG
E S S V P I R A I R A T L I L L P L L G L H Y I S T P F R P G F N S A F Y

AGGCATATGAAATTTATTCGGCGGTGACAACCTCCTTCCAGGGTCTAGCTGTGGCAATTCTGTTCTGTCTATGCAATCAAGAAGTCGTAGCCCAGGTTCGGCGCAAAATTC
E A Y E I Y S A V T T S F Q G L A V A I L F C L C N Q E V V A Q V R R K I
AGCTAACGTCGTGCTTTAAGCATCGCTACATGCAGTGCAGCACCAACGGGCAACCGGCTTGCGTGCAGGCGTCCGTGCATCATGACCAGCATCGAGCTAGCACTGTTGGTA
Q L T S C F K H R Y M Q C S T N G Q P A C V Q A S V H H D Q H R A S T V G
GTATACTG**TAA**GTCACTCGTACGACCGTCCGGCAGAAATAGAAGCAATGTTGTCAGCCGAAAGCGAACTCGATTCGATACGGATATGTTTTTAATTAGTTTAGTGATTTAA
S I L *****
TGACTTGACGAGTAGAAACAACATTATTAACAGATGCTGACCGAAATCCTGATACAAAAAGTATTTGTTCCTGGTATTCCGAGGATTCGTATCAGCGTACGGACATAACGT
ACACTGACTAATGGCATCAACGGAATAGAAGAGTTTCGGCGTAACCCAAAAATGATCTTTTGTTACAATACACCAACATGCTGTATATGATAAATTATATAATTCAGAAAA
AAACAAACCTTGTTTCATTGTATGTACGCGCTAGAGTGACAAATACATTGTAAATGTACATAGATAGATTGAGAGATGTATGTAGATTAGAAAAGATGATTCGTTTATGTT



B) Multiple sequence alignment:



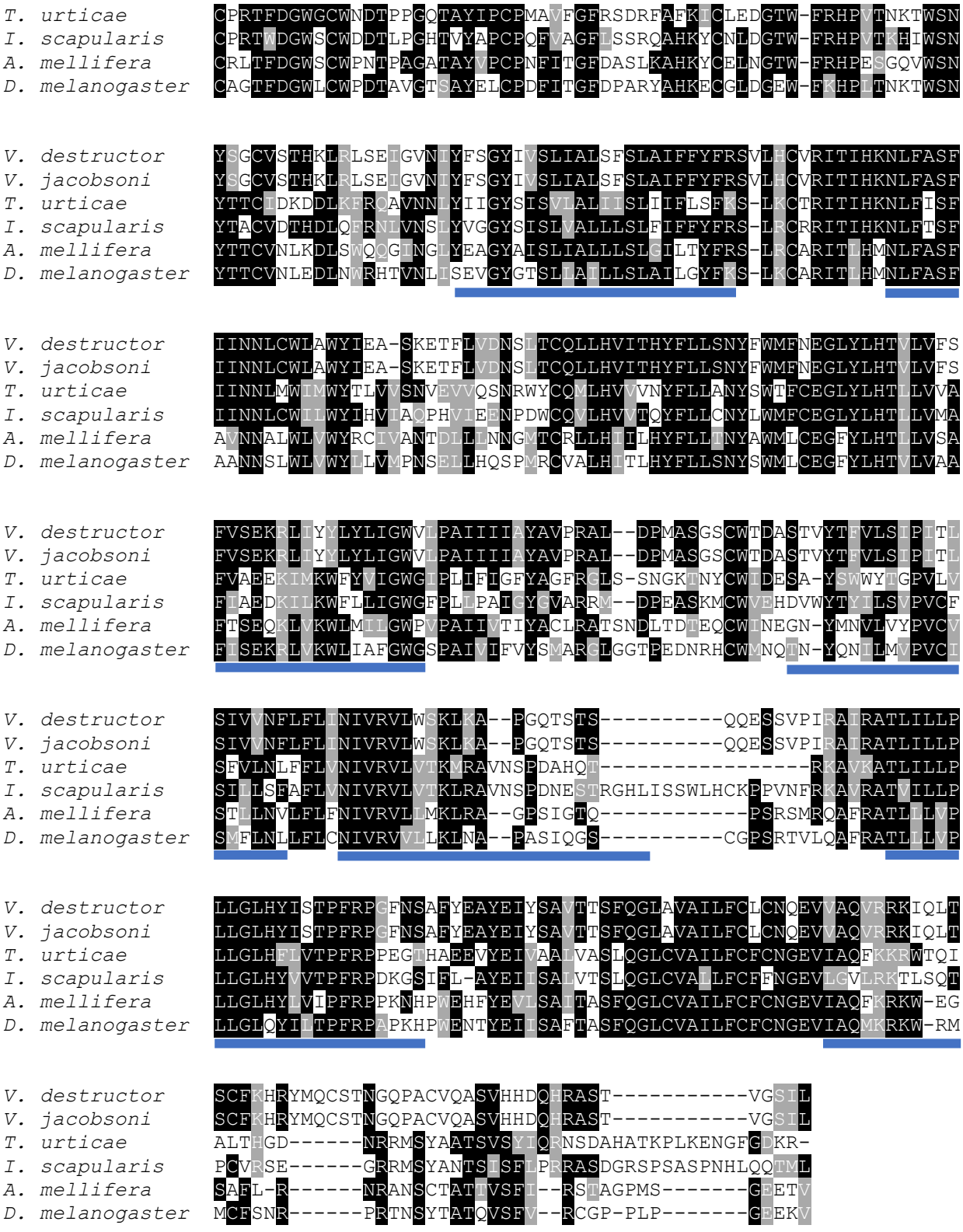


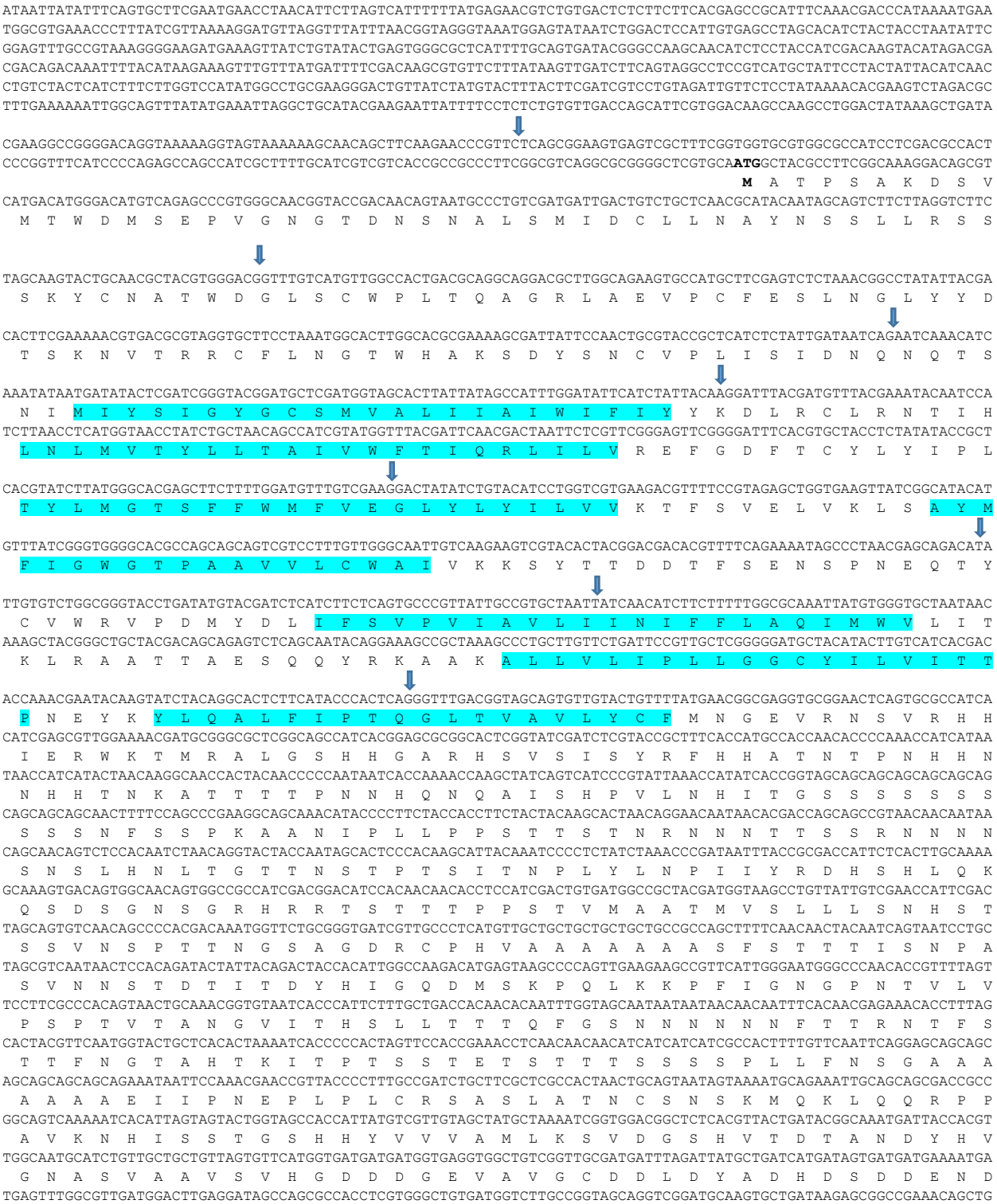
Figure 3.22 A) Gene structure for DH 31 receptor sequence in V. *destructor* and V. *jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting

and underlining. B) Multiple sequence alignment of DH31 receptor from various orthologous species across Phylum Arthropoda.

Diuretic hormone 44 type receptor

A) Gene structure:

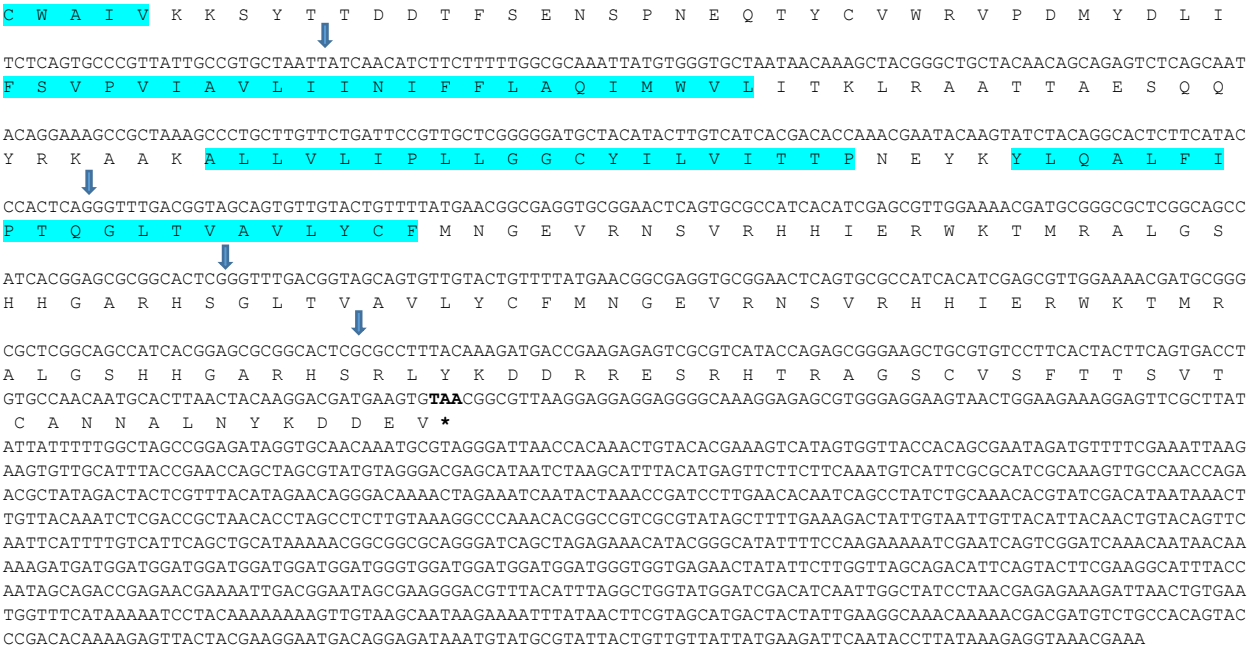
V. destructor



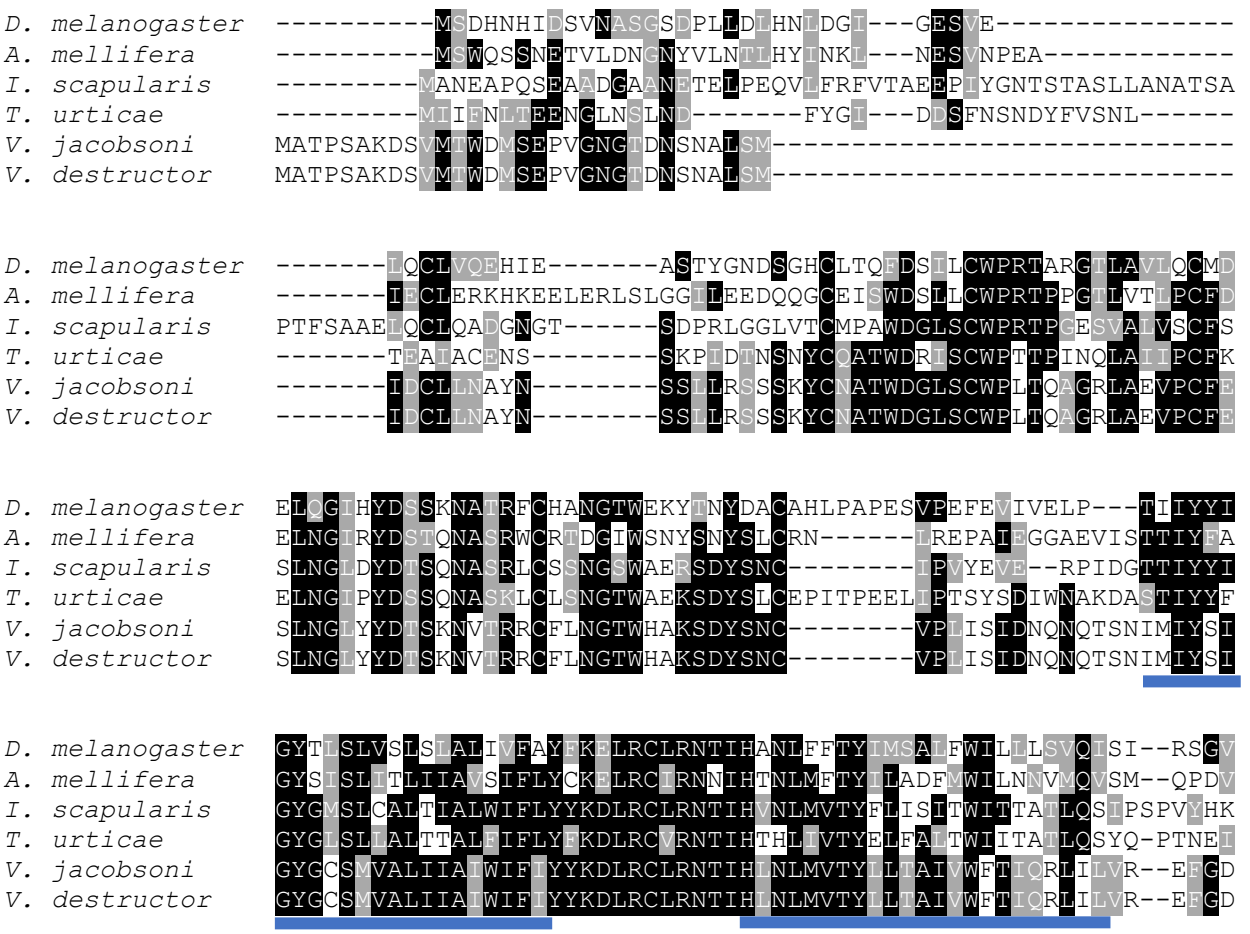
```
  E  F  G  V  D  G  L  E  D  S  Q  R  H  L  V  G  C  D  G  L  A  G  S  R  S  D  A  S  A  D  K  S  G  R  N  S  C
CAGCGGCAAGGCCAGCTGGCCACCAGGGTTAGGAGGTGTGGGGGTTGGCGGCGGCGGTGGTAGTGATTACGGGAACGAGGGCGGCGTCGGAGCCCCCTGCGCTACGCTGGC
  S  G  K  A  S  W  P  P  G  L  G  G  V  G  V  G  G  G  G  G  S  D  Y  G  N  E  G  G  V  G  A  P  C  A  T  L  A
AGGTGGTGCTGAGGCCGGAGTGCCGGGCAGTGCGCGTTATGGCCTGCTTAGCTCTCTTGGCTCGGTCGGTTCCGATATCGGCGAGCTGGGCGATCTGGCCTGTTGTAGATC
  G  G  A  E  A  G  V  P  G  S  A  R  Y  G  L  L  S  S  L  G  S  V  G  S  D  I  G  E  L  G  D  L  A  C  C  R  S
GCCTCTCTACGGTTTATCACCGGTATCGTTACCTGAACAAACTGGCTTTCTACTCCCGAGTTCCGTTCATATTGCTAATAAGGTTCACGCACCCATCGGCGTCAGTTTGGC
  P  L  Y  G  L  S  P  V  S  L  P  E  Q  T  G  F  L  L  P  S  S  V  H  I  A  N  K  V  H  A  P  I  G  V  S  L  A
CACCATCGACGTTCCTGCGGTCGCCTTGGTTACCACTACCACTAACAACAATATCAGTCATAACGAGAACAACCGTCAACTACTTCACAGGACTTCCGAACGAAGACGG TA
  T  I  D  V  P  A  V  A  L  V  T  T  T  T  N  N  N  I  S  H  N  E  N  N  R  Q  L  L  H  R  T  S  E  R  R  R  *
ATCA

CCATCACCTCGCCTAAGCGGTGTGGCCAACGTAGAGGCAACTTATAAGGCAGAAAAGTTTGCGTCCGCCATTGCCAGCGTGTGGGTTAGCGCTTCGATCGACTGGGAGTAC
GATCCTGCACTTGTCTCCGCCGGATGACGCCCGCGAACCATGCGAATGTCGCTCACACAGCAAAAGAAGTTCCGCACAGCCGGTCGTCGTCCTGCGTATCTCATATAAATG
CGTGTTCACGAGAGAAACGTCGAAATGCTCCAAAAATTACTGGAAGTGTATTATTTAGAAGATGCAATAAGCTCACATTTGTCATTTCGGATATTACAACTCATATTAAAA
GCAGGCAGAAGAGTAACGCTTATAGATCACTAACTATTTATGCCAAACAAATAATTTGAAGTCGTGCTTACGATTGGATTCTTCACTTCTGCAATCTGAAACATTTCTATA
ATTTTTCGGTGCAATTATTCGATGTCCCCGTATCTGTTCTCTTGTCGCTATTTGATTGCCAGCGCTGGATAGGGCCGCCGGGATCCCCGGATCAATCTCTCGTCTGAATAC
CGGTAACTGATTGAGATCTCGTGTGCATGGAAAGTGCAACGAGCTGTTTGTGCGTAAAGGTTTCATTCGGGTGACATGTGCGGCTGTTCTATATGTCAATCGTGTATCTGT
TCGTGACGCCACGTATTCACTTTACCGCGTACGACCTCACCGATTTAGTAAATGTTGTGTGAAGCCGATCCCGACCCGCCCACGAAGCAAGCACACTGCGTTCGTCCTTGC
ATATTAGAAGGTTCCCAAGCCAGACTGAATGTCAAGCAGATGCATCAATCACAGGATGAAAAAGAACTCGATGACGATACACGAAGCTAGAGATGTATTGCCACCACGTCA
CACCCTTCCGCATGCTATTTACTGCATATGTCTGAGAATGCGCTGAAAGTAAAAGGAGTAACCCTGGAGGCCATGGAGAGATATATCGCTTGCCTGCAATGAGCCTGATGC
TATGTATGACTGGTTCCATTTCAGTTATTTCGCAACTGATTCGCTCGTTAGTGAATTATTGCATAAATAACTGTTCGGTTGTCGGCACGCTGATCCGACCTTCAGCTATGC
ATCAGCATATAAAATGCCTGAACTGAAAATCAGAGAACTGAAAGTATGAAAACGATCATGGAGGACATGAAGATGTGTAGGCATAAAGGAGTTTAATTAACGACTACATGT
TCATTCTTGTCACCCTATGTAACGGTGATGCAGAGTACGTCTATTTGCGCTGGCCCGTAATCTTGGCATCGCTTCACTTCAATTCACTTCACTTACTTATGCCCCTGTTTG
CGTCTTTGTAAATGCCTGTTATTCGTGTTTTTGTGCTTTTTTTATGTCAAAAAGTCACGAAGCTCGAAGCCCCTTCCTCCAGAATTATCCGCGCGATACCCTACGGTTAAC
CCGTATCAGTCGTGGTCTCCCCTCGTGCTGTCTGTCCGACAAGCAAGGCCTGCATATAACGCATGCATCCTAGGGGCCAAATGAAGTATATTCGGCCAACAATGACCCTTG
AGGAGCAAAAGAACAAACTGGCAAGTGAATTATCGATAAATCTAACATGTACCATTTTTTTCTGTGTTGTCTTATCTCTATCCCCTAACACAACGTAACGTTTGTGTCACC
GCTTGCATTGCTGCTTCTCATCCGGATTACATCGCGTTACTGGGACGGGATTTTATTCCCCGTTAACTTCAACTTCGGCTAATGTTGAGTTGCGTTACATATGTATATTGC
AATGCACCACGTTGCTGCCGCTGCAGCGCCTTTACAAAGATGACCGAAGAGAGTCGCGTCATACCAGAGCGGGAAGCTGCGTGTCCTTCACTACTTCAGTGACCTGTGCCA
ACAATGCAC
```

V. jacobsoni

```
TTACGGCCTCACCAACGTCAACAACTGTTACTAACACAACTCGCGGCAATAACAGTGCCAACGTAGAAGACTATTATGTTGTAACGACTTGTCGGAAACGAACGCTCAAGC
GCGGGTTCGCGCGCAACGAACTTGACCCATTCAAAACAACTGATGCTCGACGCCGTCACAGATCACCAAATCAATCAAAGTCGATCGCCATCAGCCTCCGATTGACGCGTT
GTATATAGAATACAACAGAGTTCAGTGGTGCTGGCAGCGAAGACAAGCTTAAAACAAAAGCTTTGTTATTGTTTTCCTTTCTAATAATGCGAGGGTTAATTGGTAAAAAGG
TCAGACCCGTAGGGAATTGCAACAATCTTTGTGATTATATCAGTATGGTATAGCCACAGTGGAAGCTCGGCAGTGCTAAAGAAAGAAATGATTTTCGGCGCGTCCTGAAGC
TTCAGAGTCACCACAAATTGAAATTCGACATCCCGCTATAATCGGCTACCACTAGTAGTAGTGGCGAAATTGTCTGTCAAGCTGGTTTAGCGTATCTGTTTCACGGAACGG
CGATCCCCTCGAGATCTTTTCCACTTCTATCTTAATCTACGCACTCACTTGAGTGCGGTGCGGTGCTGTTTGATGGACGTTTGGGTTCATTAGCGGTGAACTGAAGTTAGC
AGATGATGGTAGTTGTAAACAAAGTTCTCGATGATTACATTACGTGCTTCAACAATTATTCGGTCAGTTAGGTGTGTGTGATGGTTATTCCCACCTATAATATCTGTGCGT
GTTACGTGCTTGCTACTTCAGGAAGACGACTTCACCCATAACGGAAATAGCCGCAACGTGCGATCTTCGTTGAATGTATTCGTAGGTCTTCATAACGGTATTACTTATAGC
TCAAATACGGCTAACAGTATACGCGCGCAAGCTCCACAAAGATCTGCATGCCACGGATACTCCTATATGACAAAATCGAAGCTCAGGCAGTGTTTTCATGACGAGGATCAC
TGTCACCAGCAGCGTACCGAACGAGGCCGCTATTACTAGTTTTAGATTATTCCATATGTTCCTACTGTATAATTGGAGACCTGATTCAAACGTGACGCCAGAAGTGAAATC
GTTCTCATTAAAGATAATATTATTAGTTAGGGCTGAATTTGTGCGATTCCTCTAGAGTCGTGGATTAATCGTAACGCGAGTGAAGGCTGCTTAAGTAAAACATCTAGCGCC
TCTGGCTGTGTCTGTGACGAATAATGTGCGTTACTGCGAAAAGGGGTCATAGACATGCCAAGCTATGTACTCTTAAGCAAACGTCTGTTCCCTATGATGACGAAGATAACA

TTTGAGGAGGTTTGCTGATACCGAAAGACGCTAGATTCATCTAAAAATGAAGTTGAGTCAGAAGGGTAGAACTATCAAAGCATCGTTATCCTAGCTATTGTCGTTAACTAG
TCTCGTAGCGACTTCTTTCTGGTGTGCGGATACCACCTCTGGGCCTTGTGGAAAAGCCACAAGTGATTGCTCCGTACGTGGTAAGCGTCGTTAGAGCCAGATAACAGCTAG
ACAAGCAGCAGCTAAAGAAACGGTGAAGCGGAAACCAAAAGTGATAAGGGGCAGAAAAAAATTATTAGGAGTTCTCCAGGGGAGGATGGGTCCCGTTTCACGACTCGGCCC
ATTCTCGCTGACGACACCCAAACAGCAGCAGTCAAGCCTTTGCCTGAAGAAGCCGTGCGTTCTCTAGGGCAGGCTTTGGCAGTCGCTTGCAACGTGACATATCCGTTTCAT

GTAAGCAACGAATACGCCAATCAGCGGAAGTGAGTCGCTTTCGGTGGTGCGTGGCGCCATCCTCGACGCCACTCCCGGTTTCATCCCCAGAGCCAGCCATCGCTTTTGCAT
CGTCGTCACCGCCGCCCTTCGGCGTCAGGCGCGGGGCTCGTGCAATGGCTACGCCTTCGGCAAAGGACAGCGTCATGACATGGGACATGTCAGAGCCCGTGGGCAACGGTA
                                      M  A  T  P  S  A  K  D  S  V  M  T  W  D  M  S  E  P  V  G  N  G

CCGACAACAGTAATGCCCTGTCGATGATTGACTGTCTGCTCAACGCATACAATAGCAGTCTTCTTAGGTCTTCTAGCAAGTACTGCAACGCTACGTGGGACGGTTTGTCAT
T  D  N  S  N  A  L  S  M  I  D  C  L  L  N  A  Y  N  S  S  L  L  R  S  S  S  K  Y  C  N  A  T  W  D  G  L  S
GTTGGCCACTGACGCAGGCAGGACGCTTGGCAGAAGTGCCATGCTTCGAGTCTCTAAACGGCCTATATTACGACACTTCGAAAAACGTGACGCGTAGGTGCTTCCTAAATG
C  W  P  L  T  Q  A  G  R  L  A  E  V  P  C  F  E  S  L  N  G  L  Y  Y  D  T  S  K  N  V  T  R  R  C  F  L  N

GCACTTGGCACGCGAAAAGCGATTATTCCAACTGCGTACCGCTCATCTCTATTGATAATCAGAATCAAACATCAAATATAATGATATACTCGATCGGGTACGGATGCTCGA
G  T  W  H  A  K  S  D  Y  S  N  C  V  P  L  I  S  I  D  N  Q  N  Q  T  S  N  I  M  I  Y  S  I  G  Y  G  C  S

TGGTAGCACTTATTATAGCCATTTGGATATTCATCTATTACAAGGATTTACGATGTTTACGAAATACAATCCATCTTAACCTCATGGTAACCTATCTGCTAACAGCCATCG
M  V  A  L  I  I  A  I  W  I  F  I  Y  Y  K  D  L  R  C  L  R  N  T  I  H  L  N  L  M  V  T  Y  L  L  T  A  I
TATGGTTTACGATTCAACGACTAATTCTCGTTCGGGAGTTCGGGGATTTCACGTGCTACCTCTATATACCGCTCACGTATCTTATGGGCACGAGCTTCTTTTGGATGTTTG
V  W  F  T  I  Q  R  L  I  L  V  R  E  F  G  D  F  T  C  Y  L  Y  I  P  L  T  Y  L  M  G  T  S  F  F  W  M  F

TCGAAGGACTATATCTGTACATCCTGGTCGTGAAGACGTTTTCCGTAGAGCTGGTGAAGTTATCGGCATACATGTTTATCGGCTGGGGCACGCCAGCAGCAGTCGTCCTTT
V  E  G  L  Y  L  Y  I  L  V  V  K  T  F  S  V  E  L  V  K  L  S  A  Y  M  F  I  G  W  G  T  P  A  A  V  V  L

GTTGGGCAATTGTCAAGAAGTCGTACACTACGGACGACACGTTTTCAGAAAATAGCCCTAACGAGCAGACATATTGTGTCTGGCGGGTACCTGATATGTACGATCTCATCT
```



B) Multiple sequence alignment:



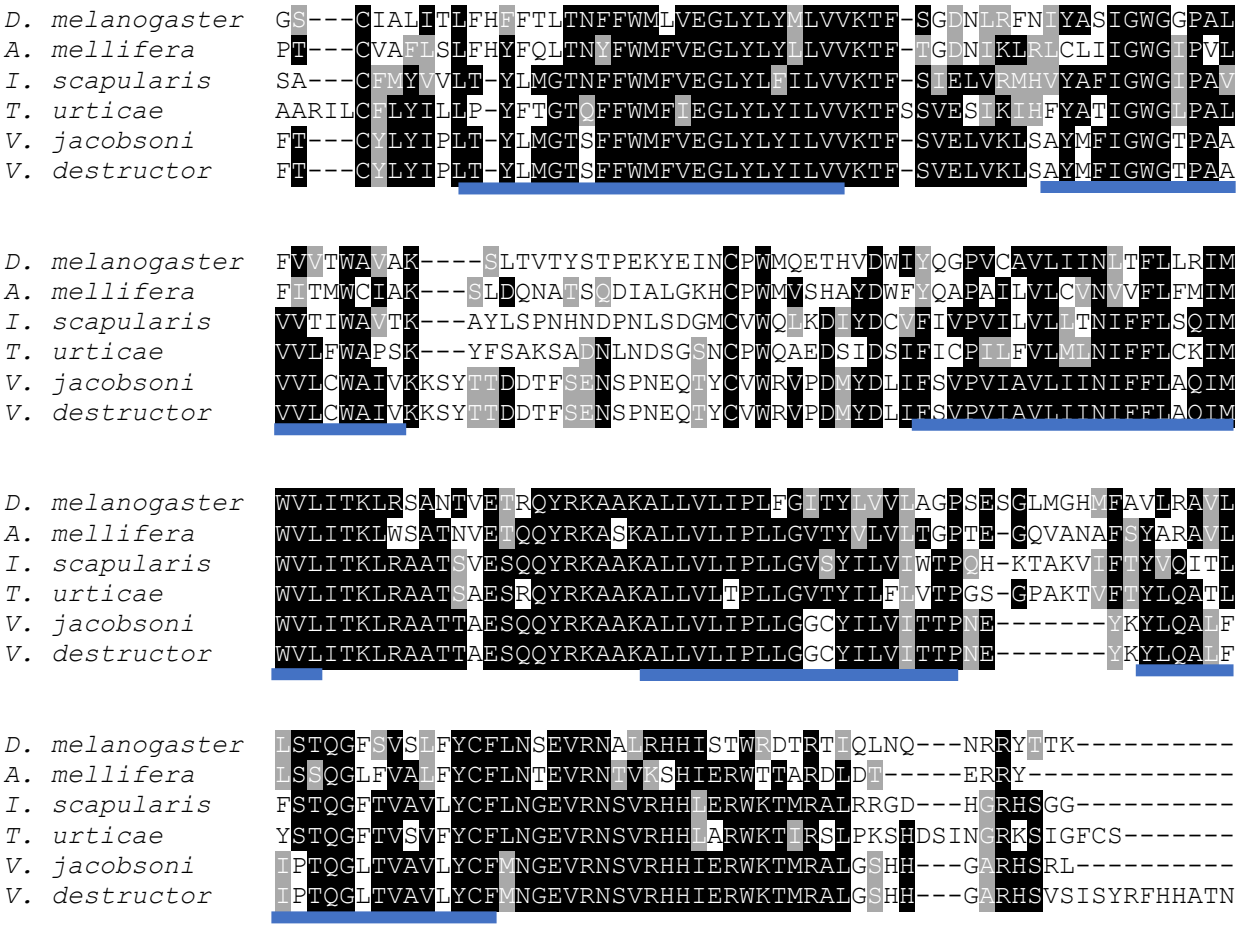


Figure 3.23 A) Gene structure for DH44 receptor sequence in V. *destructor* and V. *jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of DH44 receptor from various orthologous species across Phylum Arthropoda.

PDF receptor

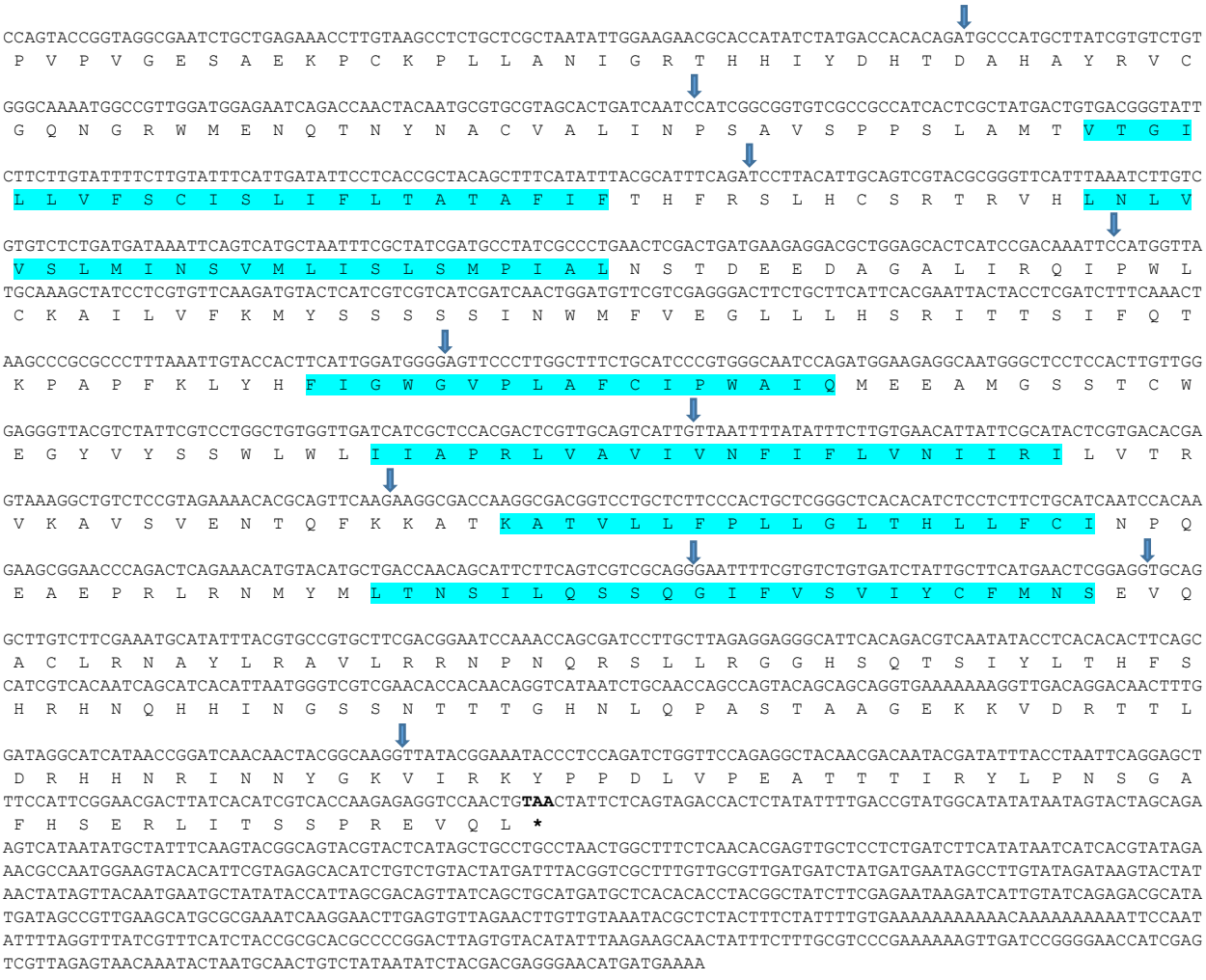
A) Gene structure

V. destructor

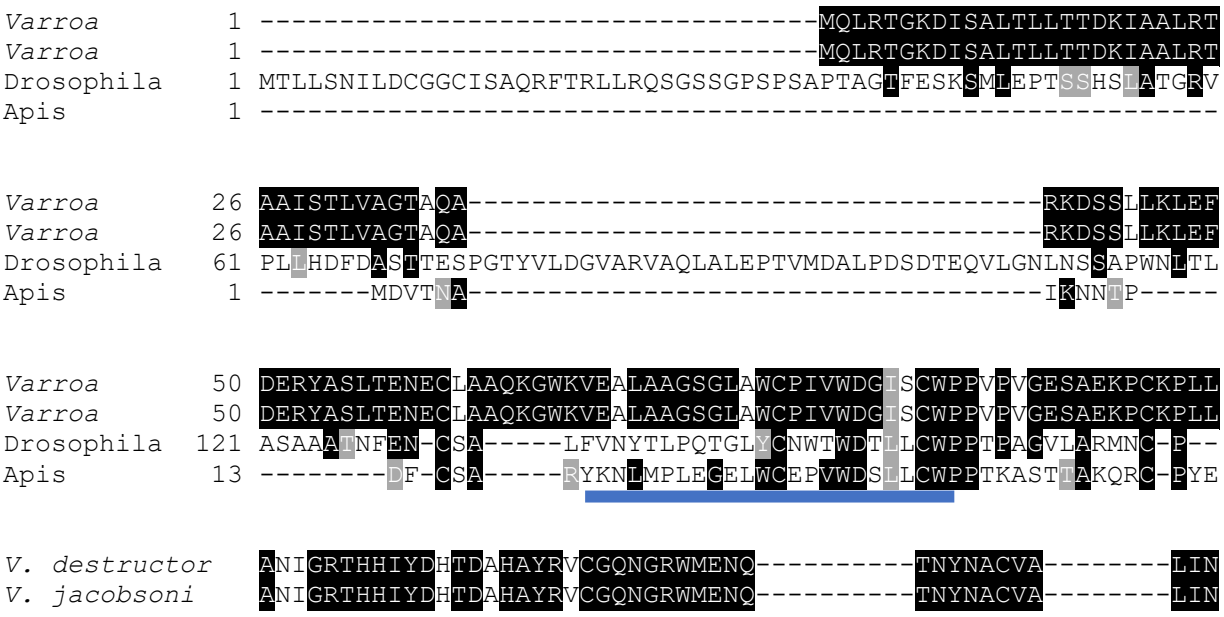
```
CGAAGAAGTCGGCACCAATAGCAGTAGAAACAGCATTAGCACGCTATGATGAAGGGATTGTGTCCGTGAAGTTCTTGTGACCAGGCGAAAAACTCTGCGGGAACAGAGCTG
ATCAGTCGTAGCTTGCTTTTCATCAAAAGCGTTCCGTGTCACGTAAATGCGCCAGCCAAAAATCAACTTCTTCAGGCCTCTGCTGTGCGACTTTGAGAAACTTTTAAGATT
AAGTGCATATGTTCAGTGTGCGCACACGTGTACAGATATTTTCCCTCTATAAAAATGAGCTTCGCTTTAAGATGTGACAACAAGTAGTGCCAGCGATTTCCAACTGTTTAA
ACCTAACAATATTGTACACTGGTGCTTTTTTATTTCCATTAAAAAGTAGATTCTTTGGTGATTGATTATCCTCGCTAGCAAACAACTAGAGGTTTTCTTAGGTCTGCGTTC
GCAAGTGATAAATCACCTACTAGTGACAATATGCAATTAAGAACGGGTAAGGATATATCTGCATTGACTCTGCTTACTACTGACAAGATTGCCGCCTTAAGGACGGCGGCC
                                  M  Q  L  R  T  G  K  D  I  S  A  L  T  L  L  T  T  D  K  I  A  A  L  R  T  A  A
ATATCAACTTTAGTAGCGGGCACTGCACAGGCGCGAAAAGATAGTAGTCTATTGAAGCTTGAATTTGATGAACGTTATGCGTCTCTAACAGAGAATGAATGTCTGGCTGCA
 I  S  T  L  V  A  G  T  A  Q  A  R  K  D  S  S  L  L  K  L  E  F  D  E  R  Y  A  S  L  T  E  N  E  C  L  A  A
                                 ⬇
CAGAAGGGCTGGAAAGTCGAAGCTTTAGCAGCTGGCTCTGGTTTAGCATGGTGTCCGATAGTCTGGGACGGCATCTCCTGCTGGCCGCCAGTACCGGTAGGCGAATCTGCT
 Q  K  G  W  K  V  E  A  L  A  A  G  S  G  L  A  W  C  P  I  V  W  D  G  I  S  C  W  P  P  V  P  V  G  E  S  A
                                                   ⬇
GAGAAACCTTGTAAGCCTCTGCTCGCTAATATTGGAAGAACGCACCATATCTATGACCACACAGATGCCCATGCTTATCGTGTCTGTGGGCAAAATGGCCGTTGGATGGAG
 E  K  P  C  K  P  L  L  A  N  I  G  R  T  H  H  I  Y  D  H  T  D  A  H  A  Y  R  V  C  G  Q  N  G  R  W  M  E
                                       ⬇
AATCAGACCAACTACAATGCGTGCGTAGCACTGATCAATCCATCGGCGGTGTCGCCGCCATCACTCGCTATGACTGTGACGGGTATTCTTCTTGTATTTTCTTGTATTTCA
 N  Q  T  N  Y  N  A  C  V  A  L  I  N  P  S  A  V  S  P  P  S  L  A  M  T  V  T  G  I  L  L  V  F  S  C  I  S
                                             ⬇
TTGATATTCCTCACCGCTACAGCTTTCATATTTACGCATTTCAGATCCTTACATTGCAGTCGTACGCGGGTTCATTTAAATCTTGTCGTGTCTCTGATGATAAATTCAGTC
 L  I  F  L  T  A  T  A  F  I  F  T  H  F  R  S  L  H  C  S  R  T  R  V  H  L  N  L  V  V  S  L  M  I  N  S  V
                                                         ⬇
ATGCTAATTTCGCTATCGATGCCTATCGCCCTGAACTCGACTGATGAAGAGGACGCTGGAGCACTCATCCGACAAATTCCATGGTTATGCAAAGCTATCCTCGTGTTCAAG
 M  L  I  S  L  S  M  P  I  A  L  N  S  T  D  E  E  D  A  G  A  L  I  R  Q  I  P  W  L  C  K  A  I  L  V  F  K
ATGTACTCATCGTCGTCATCGATCAACTGGATGTTCGTCGAGGGACTTCTGCTTCATTCACGAATTACTACCTCGATCTTTCAAACTAAGCCCGCGCCCTTTAAATTGTAC
 M  Y  S  S  S  S  S  I  N  W  M  F  V  E  G  L  L  L  H  S  R  I  T  T  S  I  F  Q  T  K  P  A  P  F  K  L  Y
                ⬇
CACTTCATTGGATGGGGAGTTCCCTTGGCTTTCTGCATCCCGTGGGCAATCCAGATGGAAGAGGCAATGGGCTCCTCCACTTGTTGGGAGGGTTACGTCTATTCGTCCTGG
 H  F  I  G  W  G  V  P  L  A  F  C  I  P  W  A  I  Q  M  E  E  A  M  G  S  S  T  C  W  E  G  Y  V  Y  S  S  W
                                       ⬇
CTGTGGTTGATCATCGCTCCACGACTCGTTGCAGTCATTGTTAATTTTATATTTCTTGTGAACATTATTCGCATACTCGTGACACGAGTAAAGGCTGTCTCCGTAGAAAAC
 L  W  L  I  I  A  P  R  L  V  A  V  I  V  N  F  I  F  L  V  N  I  I  R  I  L  V  T  R  V  K  A  V  S  V  E  N
          ⬇
ACGCAGTTCAAGAAGGCGACCAAGGCGACGGTCCTGCTCTTCCCACTGCTCGGGCTCACACATCTCCTCTTCTGCATCAATCCACAAGAAGCGGAACCCAGACTCAGAAAC
 T  Q  F  K  K  A  T  K  A  T  V  L  L  F  P  L  L  G  L  T  H  L  L  F  C  I  N  P  Q  E  A  E  P  R  L  R  N
                                       ⬇                                    ⬇
ATGTACATGCTGACCAACAGCATTCTTCAGTCGTCGCAGGGAATTTTCGTGTCTGTGATCTATTGCTTCATGAACTCGGAGGTGCAGGCTTGTCTTCGAAATGCATATTTA
 M  Y  M  L  T  N  S  I  L  Q  S  S  Q  G  I  F  V  S  V  I  Y  C  F  M  N  S  E  V  Q  A  C  L  R  N  A  Y  L
CGTGCCGTGCTTCGACGGAATCCAAACCAGCGATCCTTGCTTAGAGGAGGGCATTCACAGACGTCAATATACCTCACACACTTCAGCCATCGTCACAATCAGCATCACATT
 R  A  V  L  R  R  N  P  N  Q  R  S  L  L  R  G  G  H  S  Q  T  S  I  Y  L  T  H  F  S  H  R  H  N  Q  H  H  I
AATGGGTCGTCGAACACCACAACAGGTCATAATCTGCAACCAGCCAGTACAGCAGCAGGTGAAAAAAAGGTTGACAGGACAACTTTGGATAGGCATCATAACCGGATAAAC
 N  G  S  S  N  T  T  T  G  H  N  L  Q  P  A  S  T  A  A  G  E  K  K  V  D  R  T  T  L  D  R  H  H  N  R  I  N
          ⬇
AACTACGGCAAGAGACAACTTACCAGGCTCCGCGGGGCGACC TAA GCACACGACTTAGATGTAGACGGGTCGAACCGGTTTGTAATGGAACACGCCCTCACCTGGCTCCAT
 N  Y  G  K  R  Q  L  T  R  L  R  G  A  T  *

GGGGCGGTCACAAATTTAAGACTGACTTAAAACTTCACCAAAATAAAGGCGATTTGGCACCTCGTGAAACGGCCTCTTGTTTAGGTAAACCGGTATCAAAGTTGCTATTTT
CCGTGAACGGCAAAAGTCTCACTGATTAAAAGACGTAAGCAAGAGTTCAGCTTGGTCGCATGGCTCTCGGAACTGTTGTCAGCTTAAGGTAGTTATGATCACAGCACATAA
GACACGCGTTGGCTGGGAAGATCTTACTGTGAA
```

V. jacobsoni

```
ATTTAGTTCGCATCTGACGCTAGACGAAGAAGTCGGCACCAATAGCAGTAGAAACAGCATTAGCACGCTATGATGAAGGGATTGTGTCCGTGAAGTTCTTGTGACCAGGCG
AAAAACTCTGCGGGAACAGAGCTGATCAGTCGTAGCTTGCTTTTCATCAAAAGCGTTCCGTGTCACGTGAATGCGCCAGCCAAAAATCAACTTCTTCAGGCCTCTGCTGTG
CGACTTTGAGAAACTTTTAAGATTAAGTGCATATGTTCAGTGTGCGCACACGTGTACAGATATTTTCCCTCTATAAAAATGAGCTTCGCTTTAAGATGTGACAACAAGTAG
TGCCAGCGATTTCCAACTGTTTAAACCTAACAATATTGTACACTGGTGCTTTTTTATTTCCATTAAAAAGTAGATTCTTTGGTGATTGATTATCCTCGCTAGCAAACAACT
AGAGGTTTTCTTAGGTCTGCGTTCGCAAGTGATAAATCACCTACTAGTGACAAT ATG CAATTAAGAACGGGTAAGGATATATCTGCATTGACTCTGCTTACTACTGACAAG
                                                          M  Q  L  R  T  G  K  D  I  S  A  L  T  L  L  T  T  D  K
ATTGCCGCCTTAAGGACGGCGGCCATATCAACTTTAGTAGCGGGCACTGCACAGGCGCGAAAAGATAGTAGTCTATTGAAGCTTGAATTTGATGAACGTTATGCGTCTCTA
 I  A  A  L  R  T  A  A  I  S  T  L  V  A  G  T  A  Q  A  R  K  D  S  S  L  L  K  L  E  F  D  E  R  Y  A  S  L
                                                               ⬇
ACAGAGAATGAATGTCTGGCTGCACAGAAGGGCTGGAAAGTCGAAGCTTTAGCAGCTGGCTCTGGTTTAGCATGGTGTCCGATAGTCTGGGACGGCATCTCCTGCTGGCCG
 T  E  N  E  C  L  A  A  Q  K  G  W  K  V  E  A  L  A  A  G  S  G  L  A  W  C  P  I  V  W  D  G  I  S  C  W  P
```



B) Multiple sequence alignment:



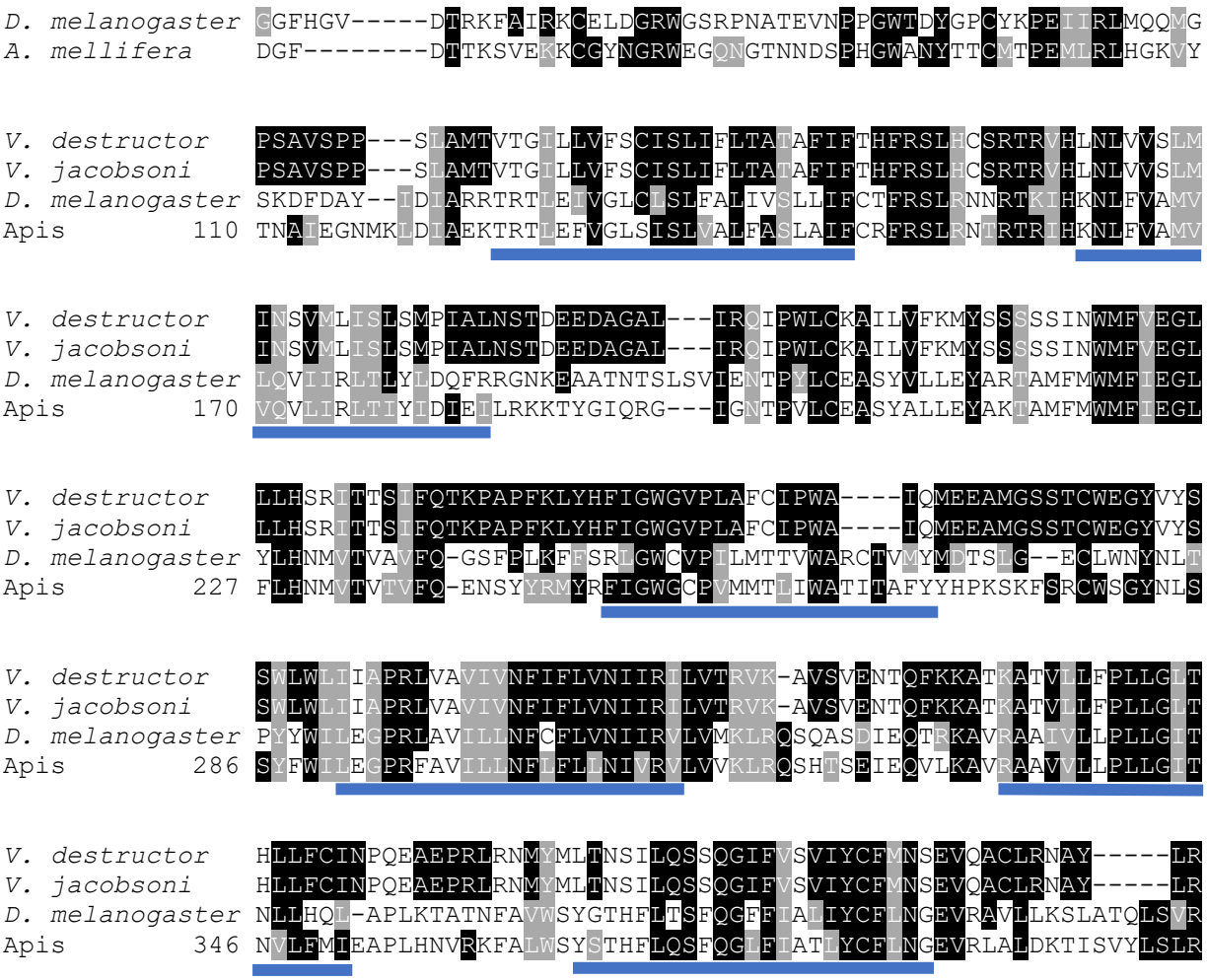


Figure 3.24 A) Gene structure for PDF receptor sequence in V. *destructor* and V. *jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of PDF receptor from various orthologous species across Phylum Arthropoda.

Insect Parathyroid hormone receptor

A) Gene structure:

V. destructor

iPTHR1

```
TCTCGCCAAGTCTGATATCCTGCCGAAACGTGCCTTCGTTAACTAACGATGACATTCAATTGCTTGGTTATTGCTTGCAAGGAACTTCATTATCCGGAAAAAAACGATGTG
TTGAAGAAAACATAGTAATTTGTTCATAGGTTCATCGTAACCGTGTAGTGGTAACGGATCAAAACCGTACAGAATGATTATGGATTATTTGGATTAGGCAGACCTTTAGGG
                    ⇩
ATCGTGTGCCCACAGCCACTTATAACTACAGGTGGTACATTTTGCAGAATTCGTCGGCATAGTACCAGCTAGATCTAACGGCTAATAGACCAAGACG ATG GAGACCGACGA
                                                                                                    ⇩      M  E  T  D  E
AGACATGATGAAGAGGTTTCTTGCTGAAGCTAAACGTAATTGCCTGTCTATCGCAAATAATGGCAACGACAGCTCTAACGAATTCAATGAAGTGACATCCTTCGATCGCTG
  D  M  M  K  R  F  L  A  E  A  K  R  N  C  L  S  I  A  N  N  G  N  D  S  S  N  E  F  N  E  V  T  S  F  D  R  C
                                                                                                          ⇩
CCCAAGAGAATGGGATGGTTTTCTGTGTTGGCCGTCTTCTCCTGCTAGCCAGAGTGTTCAGCTAAGGTGTCCCGATATGCTCTATGCCTTTGACACTTCACAGTTCGCCAG
  P  R  E  W  D  G  F  L  C  W  P  S  S  P  A  S  Q  S  V  Q  L  R  C  P  D  M  L  Y  A  F  D  T  S  Q  F  A  S


TCGTACCTGCGGAGACAATGGAACGTGGACACTGGTCAAAATAACTCCAATAGGACCTTTGTTCACGAACTACACGATGTGCACTCTTGCAACACAGGATCATATCAACAC
  R  T  C  G  D  N  G  T  W  T  L  V  K  I  T  P  I  G  P  L  F  T  N  Y  T  M  C  T  L  A  T  Q  D  H  I  N  T
           ⇩                                                                                                    ⇩
CATTGAAGCCTTTCAGCCGCATATCCGTATGATTAAATTTTTGGCCCGCATAGGATACGGAGTTTCACTCGTGAGTCTCATCTGCGCTCTGATCATCCTCATCTCTATCAA
  I  E  A  F  Q  P  H  I  R  M  I  K  F  L  A  R  I  G  Y  G  V  S  L  V  S  L  I  C  A  L  I  I  L  I  S  I  K
ACGTTTGCGGTGCCCTCGAAACTGCCTCCATATGAATCTCTTCACATCATTTATTCTACGAGCGGCCATTTTCCTTCTGAAGGATCGGATGTTTATTTCTGGAGTTGGCCT
  R  L  R  C  P  R  N  C  L  H  M  N  L  F  T  S  F  I  L  R  A  A  I  F  L  L  K  D  R  M  F  I  S  G  V  G  L
                    ⇩
CTACGGCACTTTTGATGATAACCAGAAGAGTATCTGCTGCAAGCTCTTTCTGGCAATGTTTCACTATACACTAATGGCTAATTACTGCTGGATCCTTATGGAAGGCCTCTA
  Y  G  T  F  D  D  N  Q  K  S  I  C  C  K  L  F  L  A  M  F  H  Y  T  L  M  A  N  Y  C  W  I  L  M  E  G  L  Y
                                                                                    ⇩
CTTGCATAGTCTCGTCTTTCATTCGCTTGGCAACGATCCTTCGAGTATATCGAAGTATACCATCATGGGATGGGGCTTACCTGTGCTATTTATTGCACCTTGGTCCCTGGC
  L  H  S  L  V  F  H  S  L  G  N  D  P  S  S  I  S  K  Y  T  I  M  G  W  G  L  P  V  L  F  I  A  P  W  S  L  A
                    ⇩
TCGAGCTCTATGGGAAAACAAACTATGCTGGACGACGAATAAAATTGGGTGGCATGAATGGATCATCCGTGGTCCGATTACCCTGTCAATTGTCGTTAATTTTATTCTATT
  R  A  L  W  E  N  K  L  C  W  T  T  N  K  I  G  W  H  E  W  I  I  R  G  P  I  T  L  S  I  V  V  N  F  I  L  F
                                                                                    ⇩
TGTAAACATCACAAGGGTTCTTTTTGTTAAGATGTTTGCTTCGCAGGCGCCCGTGGCCAAACGCTACAAGTACAGAAAATGGTTCAAATCGACGCTGGTTTTGGTTCCGCT
  V  N  I  T  R  V  L  F  V  K  M  F  A  S  Q  A  P  V  A  K  R  Y  K  Y  R  K  W  F  K  S  T  L  V  L  V  P  L
CTTTGGGTCCCATTATTCATTACTGTTAGTGGCTTCTATAGCTGCCGACCTGCTCTCGCCGCAAGTTGAAGTCTACTGGATGTACATAGATCAAACATTCTCGTCGTTTCA
  F  G  S  H  Y  S  L  L  L  V  A  S  I  A  A  D  L  L  S  P  Q  V  E  V  Y  W  M  Y  I  D  Q  T  F  S  S  F  Q
⇩
GGGTGTATTGGTAGCCTTGTTATATTGTTTCTTTAACACTGAAGTGAATGAGGCTATTCGATTTGTATTGCGTCGAAATCCACATTTCTTCTATATGCGCAAGTTGCTGAA
  G  V  L  V  A  L  L  Y  C  F  F  N  T  E  V  N  E  A  I  R  F  V  L  R  R  N  P  H  F  F  Y  M  R  K  L  L  N
                                                                                    ⇩
CACCAAACGCAACGTGGAAGCAGGCCAAGATGGCTCCGTACGATCTAGAGCTATGCGGCGGTCAAGGTCTATGGAGAACAGCTGCGAGAGATCAGATCTCTTCGTGCATGA
  T  K  R  N  V  E  A  G  Q  D  G  S  V  R  S  R  A  M  R  R  S  R  S  M  E  N  S  C  E  R  S  D  L  F  V  H  D
TCATTCATCGAATCCGTTTGCACGAAAACTTTCAACGACTTCATACACTGATGCTGTACTTGCCCATCCAGTTGTTGCACACCCAGGGCCTCTACCGCATCTCCAAAGCAT
  H  S  S  N  P  F  A  R  K  L  S  T  T  S  Y  T  D  A  V  L  A  H  P  V  V  A  H  P  G  P  L  P  H  L  Q  S  I
TGTTCATGGAGTGCACGTCCACGGCCTCCATGTCCATGCACAACCGGCTGGCAATAATCGTCTCCAAGTGATGCCGGAATTGCACATACTTACGGCTGAAGTCTCCACGTC
  V  H  G  V  H  V  H  G  L  H  V  H  A  Q  P  A  G  N  N  R  L  Q  V  M  P  E  L  H  I  L  T  A  E  V  S  T  S
GATGTCCAGATCGGGTTCGCCCCGCGAGCAAGCCAGCGACCCGAGGTCTATGCCGGTC TGA
  M  S  R  S  G  S  P  R  E  Q  A  S  D  P  R  S  M  P  V  *
```

iPTHR2

```
CATGCGAGTCGTCTATTTCTAACACTTGTTACGTCATCTCTATCTAATACGTATCGCTGTTGTTGTTGCCGCCGTTGTTGCCAGTAGCGGAACTAAAGCAGGTTTAATTTG
CTGTGCCCGCATAACAAAAAATTTGTTGTCTTTGATAAAGGAGGCTCTAGTGATAATATAGTAGTTGCTCTTTTAGGGACATGTGAGGTAGTCTAACTTTAAAGTCAGGCG
TGGACATACTTCGTTTCGTGCTGGCAGGGGTCTGCCACATGAAACGACGCCTATCGGTAGTACTGAACATGAAACTCAAATGATTTGGAAGACTCATGTATGGTGACGTTA
CTGTGACATTAAATTCAGAGTAAATGGAGTCTAACGACATTCCAACAAAAACAAGTGGATGACAATTTAGATAATTTGTATTTCGATCAGTGATTGACTATATGACACCTA
CGGCTCCGAATTACGATACTTGAGTTGTGGAGCAGGCGAGATTCTGTTGTTGTAGAAGTCCCTTCTCTGTCTCTTCCGTCCCCGTACCGTGAAAGCTGACGTAATAGCACG
GCTACATCGTGAAGGCGTACTCAGTTAATTTTGTAGTTCACGCATGAGTATCAGATCATAAATTGCAGTTAAGGGAAGATTAGTGCAATTGTTTGGAGTGAAAATCGTGAG
ATTCGGTACCCGGGAACGACCTGTTGGCATCTGACTAGGTCGCGCCTACTACCGCCTCTGCCGCCGCT ATG GATTGGCGTAACGCGAACGCGACGGACGCATCGCTTAACG
                                                                          M  D  W  R  N  A  N  A  T  D  A  S  L  N
ACACGGATGCCCTGTTCTCCGAATATGGATCGACCCGTACTGCTAGCTCGATTGCCGGTGTTGCGGTGACATCGTCAGTCGGAACAGGGTCATTGGGCATTGCTGCGTCGG
D  T  D  A  L  F  S  E  Y  G  S  T  R  T  A  S  S  I  A  G  V  A  V  T  S  S  V  G  T  G  S  L  G  I  A  A  S
CTGCCGAGTTATGCGTTGGAACAGCTTTGAACAGCAGCGATGTTGCCGTATCAGCAACATCTGAGGGGGTTGCCTGCGCCGCTGAATGGGATGGCCGTTTATGCTGGAAGA
A  A  E  L  C  V  G  T  A  L  N  S  S  D  V  A  V  S  A  T  S  E  G  V  A  C  A  A  E  W  D  G  R  L  C  W  K
ATGCAGCCCCCGGCGAGGTCGTCGCCATGTTCTGCCCAGAAGGACGCTTCTTTAACTCTTCTTATATGGTGGAACGGTCTTGCAGCACACACGGTGTATGGAGTATTGCAA
```

```
N A A P G E V V A M F C P E G R F F N S S Y M V E R S C S T H G V W S I A
AGTACGAAGCGTGCTTTGAACAGGCCTACTATGATTACATTCATACGTTGCAGGGTCACCTCTCGACACTAAATCGAATTTCACTTATCGGTTATTCATCGTCTCTTTTTT
K Y E A C F E Q A Y Y D Y I H T L Q G H L S T L N R I S L I G Y S S S L F

TACTGATACTGGCTTTTTTTCTTCTAACTTCTCTTAAACGTTTACGGTGTGCTCGTAACAAATTACATTTACATCTGTCTGCGTCATTCATCCTTCGATGTACTGTACTGG
L L I L A F F L L T S L K R L R C A R N K L H L H L S A S F I L R C T V L

TATTTAAGCATCTCTATTTTACCAGGGGCATAACATCCTTCAATAGTCATGAATTCATCAGTTGTCGCATCTTCACGCTTTCGTTTCATTACTCGTTGATGGCCAATTATT
V F K H L Y F T R G I T S F N S H E F I S C R I F T L S F H Y S L M A N Y

GTTGGATTTTGATGGAAGGCCTCTATCTCTACAATATCATTTTCATGTCGGTTTACGCTGATAATTCGAAAATTGCAAAATACATAGTCATGGGCTGGGGTCTTCCGCTGC
C W I L M E G L Y L Y N I I F M S V Y A D N S K I A K Y I V M G W G L P L

CTTGGGTGGCGATCTGGGCAATAGCACGGCTCTGTTACGACAATAATCGGTGTTGGGAGATGGAGCCGAAGTATCGTGGCATCACCTGGATTCTACGCGGACCCATTTCAC
P W V A I W A I A R L C Y D N N R C W E M E P K Y R G I T W I L R G P I S

TCTCTATCATTTTAAACTTTTTCTTCTTCCTAAGTATCACACGAGTTTTATTTGTCAAAATGTCAGCTGTGCACATACCGGATGCGCGAAAAGTTCGATACCGAAAATGGT
L S I I L N F F F F L S I T R V L F V K M S A V H I P D A R K V R Y R K W
TTAAATCAACACTCGTTTTAATTCCGTTATTCGGCGCTCATCAGATACTTACTTTAGCAGCCAGCCTCCTGCATAACGACCGGATGGAATTTATTTGGCTTTACATAGATG
F K S T L V L I P L F G A H Q I L T L A A S L L H N D R M E F I W L Y I D

TCGCCTTTACTACGTTTCAGGGTTCAGTGGTAGCTCTACTATACTGTTTTGTAAATAATGAGGTGCAGTCAGAGGTGATCCGACTGCTGCCTGAACGTTTTCGACGAAAAC
V A F T T F Q G S V V A L L Y C F V N N E V Q S E V I R L L P E R F R R K

TTCCCTTAGGAGACGGTGGAGGGTTCCCATCATCGACGAATTTCTCTAAACGAGATTCGCTGGCACCGGGTGTGTGTACATTCAAAGTGCACGGGAAGCGCCGTCTACTTC
L P L G D G G G F P S S T N F S K R D S L A P G V C T F K V H G K R R L L

GACAGGGATCCACACAGACGTGCACCGTTCGCGAGGAAATTAGTCTAGATAATATACAACAGTTGGGCAACATTCGGTGCCAGGATAAAAGCAATCACAAGAGCGGGCTCA
R Q G S T Q T C T V R E E I S L D N I Q Q L G N I R C Q D K S N H K S G L
GCAGCAATGGAGACCAAAAAAGCAAGAATAATTCCAATGAAACTAAGCTACTGTTAAGCAGAGGACGTCCATGCAGCGGTCGAGAAGATGTGCTCGATGGAGGTTTACATA
S S N G D Q K S K N N S N E T K L L L S R G R P C S G R E D V L D G G L H
ATGATTCATGTGTC TAA CAGAAACTCAA
N D S C V *
```

iPTHR3

```
CATTGGCGAAGAGACTTTAAAGTTACTGCAGATCGCTGGAGTTGCTGTTGTGACTGGTACGAAAGTTACCTATCGTGCGGTGACGACCCGCCGAAAGAAACACGATACAGT
TTATCTTTATGTCATGGCAAGTTTAACTTAAGAAATGACTAACAAACAAGGTGCTATGTGTTATTACTGGAATACTGAGCAAATACTCTATAGAGCGAAACGACCCTCTAA
ACCTGATGCGATAAAGTTATGTTCTATCTAATTCGACATGTGGATTTAATGAGAGCACTTTTTAGCAAACGATTTGGTATATTATAAAAGGTACACTGACTCTTATATAGT
TGATATTGTTCCTGTCGTCCATAAAGTGGTTCGTGTATTAGGATAGTGACCTTTGTGTGTGTATAGCAAAAAATGAAATTGCTTAGTGAAGTTACAGTGGAGTTAATGATA
CAATAAGGAAAACTAATCCTGGTGCCATAAACGTCATTTTGTTCTACCCATCTTGTTTGAAGCATGTCAGCGGTACATAGCGCCAATTAGGTTTTTAAGCATACCCAGTCG

CTGTAACTAGCTCCACAACAAGGAACCTTTG ATG GCATCGCACGAGGCCAATAAGGATTTGGATGAGTGGAGCAACTCAACGGTGGTGGTAGTCCCTCTACCGGAGGTTCA
                               M  A S H E A N K D L D E W S N S T V V V V P L P E V Q
GTCAATGAGCCTTTGTCTGAACTCAACGGAGCGACAAAGCAATGTAAACTTTACTCCGTATCCCATCGATGCTAACGGTGAAGGAAATAGTCTGACAGCGCGATCCTGCCC
  S M S L C L N S T E R Q S N V N F T P Y P I D A N G E G N S L T A R S C P
GCGTGACTGGGACGGCTTTCTATGTTGGCCGCAGACGCTGCCCGAGCAGATAGTTCTGCAGAGATGTCCTACCCTCAGGAGGCCAGACGCCATCGGGGTAACAGTGTCGCA
  R D W D G F L C W P Q T L P E Q I V L Q R C P T L R R P D A I G V T V S Q

GGTCGACAAGGTTGCCGAGTTCGCATGGCGACACTGCGGCGCAAACGGCACGTGGACTTCGCTGCAGTCCAGTGACTCGTATAGCTGGCACCTGTATACAAACTTCTCCGC
  V D K V A E F A W R H C G A N G T W T S L Q S S D S Y S W H L Y T N F S A

CTGCAATATGGACCCTACAGTTGCTCTATCACCGAATGAGCAAGAACTTTACAGATTAATCCGAGTGCATCTGCCTGCGATTAAGCAGATGTCGATCGTAGGATACTCCAC
  C N M D P T V A L S P N E Q E L Y R L I R V H L P A I K Q M S I V G Y S T

ATCACTAGCACTTCTACTTGGTGCCTTCACGCTACTCGCATCGCTCAAGCGTTTACGATGCGCACGCAATAAGATACACCTGCATCTTTTTGCTTCGTTTATTGTTCGGGC
  S L A L L L G A F T L L A S L K R L R C A R N K I H L H L F A S F I V R A

TGTTGTTCTGGTCCATAAGCATAGTTCAACGAGGCATTCCGATCTGACTCCACAGTGTTGGATGCTGATGTGCCTGTTTCACTATTCACTCATGGCTAATTACTGCTGGAT
  V V L V H K H S S T R H S D L T P Q C W M L M C L F H Y S L M A N Y C W I

CCTCGTCGAAGGTCTTTACCTGCATAACCTTATCTTTAACTCGTTCTACGCAGATGGATCTAGTATCGCTAAATACGTTGTCATGGGCTGGGGTTTGCCTGTGTCGTGCGT
  L V E G L Y L H N L I F N S F Y A D G S S I A K Y V V M G W G L P V S C V

AATCGTATGGGCGATTGCTCGAAATATCGTAGGCAACGACACGTGTTGGACGTCAGACGCAAAGCTGATCCAATGGCTGCTACGAGGGCCCATCACTCTTAGTATAGTGCT
  I V W A I A R N I V G N D T C W T S D A K L I Q W L L R G P I T L S I V L

CAACTTTATTTTCTTTATCAATATTACTCGCGTTGTGTTCATCAAGATGTCGAATTCGCACGAGCCTGAAGCGCGGCGCGTGAAATACCGCAAATGGTTCAAGTCGACCTT
  N F I F F I N I T R V V F I K M S N S H E P E A R R V K Y R K W F K S T L

AGTCTTAGTTCCATTGTTTGGAGCGCATCACATAATCCTTATGGTGATGTCAATTGCAGCCGTTACGCCGCTGTACGAACTTTACTGGCTTTACATTGATCAGCTCTTCAC
```

```
  V  L  V  P  L  F  G  A  H  H  I  I  L  M  V  M  S  I  A  A  V  T  P  L  Y  E  L  Y  W  L  Y  I  D  Q  L  F  T
ATCGTTTCAGTTCCTCAAGGGTCCTGATCCGGTGATTTTGTCCAACTGTCCAGCAAAAAGGCATAGCTGTCCGTCGATGTCCATTACCTAAACGAGTGCACAACACTGTAT
  S  F  Q  F  L  K  G  P  D  P  V  I  L  S  N  C  P  A  K  R  H  S  C  P  S  M  S  I  T  *
ATCCGCACAGTGGCTCCTCCCATTGTATTCATAAAGGTCCTGGATGGAGCTCGTTGAAACCACCTGCCGGCAACAGAAACTGGACGGTCAGCCGTTGCAAGCCGCTCTGTC
GTCAGGGCCGTTATCTATACATGAATGTTCCCGATCATTTCAGACGGAATCGAG
```

V. jacobsoni

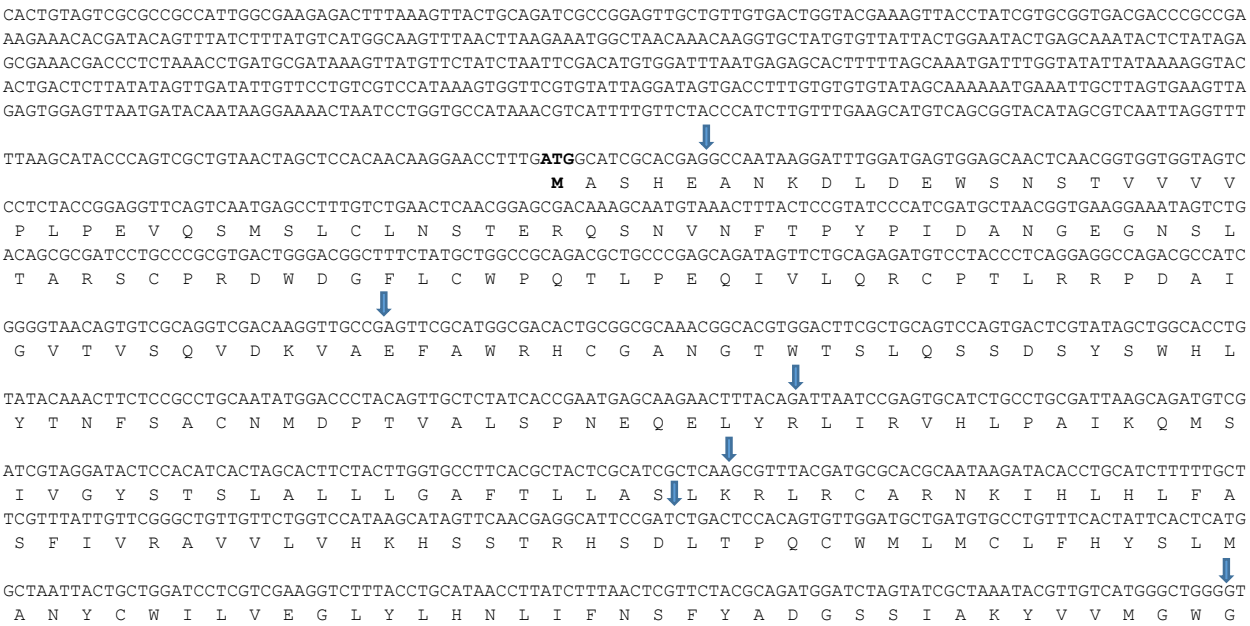
iPTHR1

```
GTGGTACATTTTGCAGAATTCGTCGGCATAGTACCAGCTAGATCTAACGGCTAATAGACCAAGACGATGGAGACCGACGAAGACATGATGAAGAGGTTTCTTGCTGAAGCT
                                                        M  E  T  D  E  D  M  M  K  R  F  L  A  E  A
AAACGTAATTGCCTGTCTATCGCAAATAATGGCAACGACAGCTCTAACGAAATAAACGGGGTTTTCAATGAAGTGACATCCTTCGATCGCTGCCCAAGAGAATGGGATGGT
 K  R  N  C  L  S  I  A  N  N  G  N  D  S  S  N  E  I  N  G  V  F  N  E  V  T  S  F  D  R  C  P  R  E  W  D  G

TTTCTGTGTTGGCCGTCTTCTCCTGCTAGCCAGAGTGTTCAGCTAAGGTGTCCCGATATGCTCTATGCCTTTGACACTTCACAGTTCGCCAGTCGTACCTGCGGAGACAAT
 F  L  C  W  P  S  S  P  A  S  Q  S  V  Q  L  R  C  P  D  M  L  Y  A  F  D  T  S  Q  F  A  S  R  T  C  G  D  N

GGAACGTGGACACTGGTCAAAATAACTCCAATAGGACCTTTGTTCACGAACTACACGATGTGCACTCTTGCAACACAGGATCATATCAACACCATTGAAGCCTTTCAGCCG
 G  T  W  T  L  V  K  I  T  P  I  G  P  L  F  T  N  Y  T  M  C  T  L  A  T  Q  D  H  I  N  T  I  E  A  F  Q  P

CATATCCGTATGATTAAATTTTTGGCCCGCATAGGATACGGAGTTTCACTCGTGAGTCTCATCTGCGCTCTGATCATCCTCATCTCTATCAAACGTTTGCGGTGCCCTCGA
 H  I  R  M  I  K  F  L  A  R  I  G  Y  G  V  S  L  V  S  L  I  C  A  L  I  I  L  I  S  I  K  R  L  R  C  P  R
AACTGCCTCCATATGAATCTCTTCACATCATTTATTCTACGAGCGGCCATTTTCCTTCTGAAGGATCGGATGTTTATTTCTGGAGTTGGCCTCTACGGCACTTTTGATGAT
 N  C  L  H  M  N  L  F  T  S  F  I  L  R  A  A  I  F  L  L  K  D  R  M  F  I  S  G  V  G  L  Y  G  T  F  D  D

AACCAGAAGAGTATCTGCTGCAAGCTCTTTCTGGCAATGTTTCACTATACACTAATGGCTAATTACTGCTGGATCCTTATGGAAGGCCTCTACTTGCATAGTCTCGTCTTT
 N  Q  K  S  I  C  C  K  L  F  L  A  M  F  H  Y  T  L  M  A  N  Y  C  W  I  L  M  E  G  L  Y  L  H  S  L  V  F

CATTCGCTTGGCAACGATCCTTCGAGTATATCGAAGTATACCATCATGGGATGGGGCTTACCTGTGCTATTTATTGCACCTTGGTCCCTGGCTCGAGCTCTATGGGAAAAC
 H  S  L  G  N  D  P  S  S  I  S  K  Y  T  I  M  G  W  G  L  P  V  L  F  I  A  P  W  S  L  A  R  A  L  W  E  N

AAACTATGCTGGACGACAAATAAAATTGGATGGCATGAATGGATCATCCGTGGTCCGATTACCCTGTCAATTGTCGTTAATTTTATTCTATTTGTAAACATCACAAGGGTT
 K  L  C  W  T  T  N  K  I  G  W  H  E  W  I  I  R  G  P  I  T  L  S  I  V  V  N  F  I  L  F  V  N  I  T  R  V

CTTTTTGTTAAGATGTTCGCTTCGCAGGCGCCCGTGGCCAAACGCTACAAGTACAGAAAATGGTTCAAATCGACGCTGGTTTTGGTTCCGCTCTTTGGGTCCCATTATTCA
 L  F  V  K  M  F  A  S  Q  A  P  V  A  K  R  Y  K  Y  R  K  W  F  K  S  T  L  V  L  V  P  L  F  G  S  H  Y  S
TTACTGTTAGTGGCTTCTATAGCTGCCGACCTGCTCTCGCCGCAAGTTGAAGTCTACTGGATGTACATAGATCAAACATTCTCGTCGTTTCAGGTAAATTTCCTTATCTTT
 L  L  L  V  A  S  I  A  A  D  L  L  S  P  Q  V  E  V  Y  W  M  Y  I  D  Q  T  F  S  S  F  Q  V  N  F  L  I  F
ATATATATTTTTTTTTAG
 I  Y  I  F  F  *
```

iPTHR2

```
TCCACCCCATCTATCGACTTAGCTCGAGCAGAAAGATGTCGAATACATGCGAGTCGTCTATTTCTAACACTTGTTACGTCATCTCTATCTAATACGTATCGCTGTTGTTGT
TGCCGCCGTTGTTGCCAGTAGCGGAACTAAAGCAGGTTTAATTTGCTGTGCCCGCATAACAAAAAAATTGTTGTCTTTGATAAAGGAGGCTCTAGTGATAATATAGTAGTT
GCTCTTTTAGGGACATGTGAGGTAGTCTAACTTTAAAGTCAGGCCTGGACATACTTCGTTTCGTGCTGGCAGGGGTCTGCCACATGAAACGACGCCTATCGGTAGTACTGA
ACATGAAACTCAAATGATTTGGAAGACTCATGTATGGTGACGTTACTGTGACATTAAATTCAGAGTAAATGGAGTCTAACGACATTCCAACAAAAACAAGTGGATGACAAT
TTAGATAATTTGTATTTCGATCAGTGATTGACTATATGACACCTACGGCTCCGAATTACGATACTTGAGTTGTGGAGCAGGCGAGATTCTGTTGTTGTAGAAGTCCCTTCT
CTGTCTCTTCCGTCCCCGTACCGTGAAAGCTGACGTGATAGCACGGCTACATCGTGAAGGCGTACTCAGTTAATTTTGTAGTTCACGCATGAGTATCAGATCATAAATTGC
AGTTAAGGGAAGATTAGTGCAATTGTTTGGAGTGAAAATCGTGAGATTCGGCACCCGGGAACGACCTGTTGGCATCTGACTAGGTCGCGCCTACTACCGCCTCTGCCGCCG
CTATGGATTGGCGTAACGCGAACGCGACGGACGCATCGCTTAACGACACGGATGCCCTGTTCTCCGAATATGGATCGACCCGTACTGCTAGCTCGATTGCCGGTGTTGCGG
   M  D  W  R  N  A  N  A  T  D  A  S  L  N  D  T  D  A  L  F  S  E  Y  G  S  T  R  T  A  S  S  I  A  G  V  A
TGACATCGTCAGTCGGAACAGGGTCATTGGGCATTGCTGCGTCGGCTGCCGAGTTATGCGTTGGAACAGCTTTGAACAGCAGCGATGTTGCCGTATCAGCAACATCTGAGG
V  T  S  S  V  G  T  G  S  L  G  I  A  A  S  A  A  E  L  C  V  G  T  A  L  N  S  S  D  V  A  V  S  A  T  S  E

GGGTTGCCTGCGCCGCTGAATGGGATGGCCGTTTATGCTGGAAGAATGCAGCCCCCGGCGAGGTCGTCGCCATGTTCTGCCCAGAAGGACGCTTCTTTAACTCTTCTTATA
G  V  A  C  A  A  E  W  D  G  R  L  C  W  K  N  A  A  P  G  E  V  V  A  M  F  C  P  E  G  R  F  F  N  S  S  Y
TGGTGGAACGGTCCTGCAGCACACACGGTGTATGGAGTATTGCAAAGTACGAAGCGTGCTTTGAACAGGCCTACTATGATTACATTCATACGTTGCAGGGTCACCTCTCGA
M  V  E  R  S  C  S  T  H  G  V  W  S  I  A  K  Y  E  A  C  F  E  Q  A  Y  Y  D  Y  I  H  T  L  Q  G  H  L  S

CACTAAATCGAATTTCACTTATCGGTTATTCATCGTCTCTTTTTTTTACTGATACTGGCTTTTTTTCTTCTAACTTCTCTTAAACGTTTACGGTGTGCTCGTAACAAATTAC
T  L  N  R  I  S  L  I  G  Y  S  S  S  L  F  L  L  I  L  A  F  F  L  L  T  S  L  K  R  L  R  C  A  R  N  K  L
ATTTACATCTGTCTGCGTCATTCATCCTTCGATGTACTGTACTGGTATTTAAGCATCTCTATTTTACCAGGGGCATAACATCCTTCAATAGTCATGAATTCATCAGTTGTC
H  L  H  L  S  A  S  F  I  L  R  C  T  V  L  V  F  K  H  L  Y  F  T  R  G  I  T  S  F  N  S  H  E  F  I  S  C
GCATCTTCACGCTTTCGTTTCATTACTCGTTGATGGCCAATTATTGTTGGATTTTGATGGAAGGCCTCTATCTCTACAATATCATTTTCATGTCGGTTTACGCTGATAATT
R  I  F  T  L  S  F  H  Y  S  L  M  A  N  Y  C  W  I  L  M  E  G  L  Y  L  Y  N  I  I  F  M  S  V  Y  A  D  N
```

iPTHR3



```
                                                                                                                    ⇓
TTGCCTGTGTCGTGCGTAATCGTATGGGCGATTGCTCGAAATATCGTAGGCAACGACACGTGTTGGACGTCAGACGCAAAGCTGATCCAATGGCTGCTACGAGGGCCCATC
 L  P  V  S  C  V  I  V  W  A  I  A  R  N  I  V  G  N  D  T  C  W  T  S  D  A  K  L  I  Q  W  L  L  R  G  P  I
          ⇓                                                                                                   ⇓
ACTCTTAGTATAGTGCTCAACTTTATTTTCTTTATCAATATTACTCGCGTTGTGTTCATCAAGATGTCGAATTCGCACGAGCCTGAAGCGCGGCGCGTGAAATACCGCAAA
 T  L  S  I  V  L  N  F  I  F  F  I  N  I  T  R  V  V  F  I  K  M  S  N  S  H  E  P  E  A  R  R  V  K  Y  R  K
TGGTTCAAGTCGACCTTAGTCTTAGTTCCATTGTTTGGAGCGCATCACATAATCCTTATGGTGATGTCAATTGCAGCCGTTACGCCGCTGTACGAACTTTACTGGCTTTAC
 W  F  K  S  T  L  V  L  V  P  L  F  G  A  H  H  I  I  L  M  V  M  S  I  A  A  V  T  P  L  Y  E  L  Y  W  L  Y

ATTGATCAGCTCTTCACATCGTTTCAGTTCCTCAAGGGTCCTGATCCGGTGATTTTGTCCAACTGTCCAGCAAAAAGGCATAGCTGTCCGTCGATGTCCATTACCTAAACG
 I  D  Q  L  F  T  S  F  Q  F  L  K  G  P  D  P  V  I  L  S  N  C  P  A  K  R  H  S  C  P  S  M  S  I  T  *
AGTGCACAACACTGTATATCCGCACAGTGGCTCCTCCCATTGTATTCATAAAGGTCCTGGATGGAGCTCGTTGAAACCACCTGCCGGCAACAGAAACTGGACGGTCAGCCG
TTGCAAGCCGCTCTGTCGTCAGGGCCGTTATCTATACATGAATGTTCCCGATCATTTCAGACGGAATAGAGGTCAGCCAGAACAGTAGGGTACAACTCTTGGACAACACCA
TAAGCGTCTAATCCGTTATCCTTGCGAATTATATATTACAGAACACGCTCTAGGTATTTAATCGGCTAGTTTCTTGAGAATTCGCTGTGAAGCTGATTTTTAGCTGTGACA
GTTTTTGTACTTCACCAGAGTTAATTTAAGAGACTTACCTTGTTAGGGCTTTACTTCACGCAAGAATGGAATAGAGCATGAAATTGCCTCGTCCAACGTACATAAAGGGTG
TACAGGCTAATATCAAAGTTTGATATTTAGTAACTCTTTTATTTCTCTATCTTTTCAACTACACACGACATCGGAGGTTAACACAAAACCTCACTTATTTTTGGGGGCATT
ATCTTTGTATAGTTACATAATGTGATCTTGTCCTCCCCGGATATTTGAAGGCAAATATGTTTGCCAAAGTTTACATTCAACAGAAACGCTCGTAGTTCAATAACTAGCGAA
GACAGCTAATGGTTTATAGAGAGCCCCTACGTAACAGCATAGGAGGCAGTAGTGCAATACTCAGCTTACAGTCGTCTAGTTTCGGTTCTGTTTTTTAGTTTCAACTCAAGT
GTACTACCTGTAGAAGTGTTTGAAAGCCAGCACGCACTCTTAATCAGAGGCAAGACGTTTAAGATTCTAAGATACAATTGGAACAGAACAAATGCAAATATGGTTATCTGG
AAAGGGAGGTAAAGGCCCGACCGGACCCATTTTATTGAAGCTAAATTGATGCGGGCATTTCCACCGTGGAAATTCATGTATAACAAGTGTCTGTACTGAATTTCAATAAGA
TATACGATGCTTCGAGTCTGGTAAAGCCAGCAGTTGTTCCCATTAGCCTTGCGAATGGGTGATTAATAAAGGGCTTTTTTTATTATTGGCGAACAGCCATTCACAACTTTA
GTACAGATTGAATTACTACGCC
```

B) Multiple sequence alignment:

```
V. jacobsoni      LPAIKQMSIVGYSTSLALLLGAFTLLASLKRLRCARNKIHLHLFASFIVRAVVLVHKHS-
V. jacobsoni 2    LSTLNRISLIGYSSSLFLLILAFFLLTSLKRLRCARNKLHLHLSASFILRCTVLVFKHLY
V. jacobsoni 3    IRMIKFLARIGYGVSLVSLICALIILISIKRLRCPRNCLHMNLFTSFILRAAIFLLKDRM
V. destructor     LPAIKQMSIVGYSTSLALLLGAFTLLASLKRLRCARNKIHLHLFASFIVRAVVLVHKHS-
V. destructor 2   IRMIKFLARIGYGVSLVSLICALIILISIKRLRCPRNCLHMNLFTSFILRAAIFLLKDRM
V. destructor 3   LSTLNRISLIGYSSSLFLLILAFFLLTSLKRLRCARNKLHLHLSASFILRCTVLVFKHLY
I. scapularis     LPTIKLISKVCYTVSLVTLLAAFLILASIKRLRCPRNSLHMHLFISFILRALAFLLKDAL
T. urticae        LPHIKLISKIGYSVSLLTLIIAFLILFCNKRLQCPRNHLHLQLFLSFISRSLLTHVKNIF
T. castaneum      LPIIKNITQCGYILSTVSLIISLFVFIRIKRLHCARNKLHIHLFASFVMRALMSLIKDGL
T. castaneum 2    APVVQVISETGYIVSFATLIIAFAIMLFIKKLHCARNILHMHLFASFILRALTFIVIKST
A. mellifera      IPIVKIISKIGYTVSFFTLVIAFFILTVIKKLRCPRNILHMHLFASFMFRAFMALMKDIV

V. jacobsoni      ---------ST-----------RHSDLTPQCWMLMCLFHYSLMANYCWILVEGLYLHNLI
V. jacobsoni 2    FTRGIT---SF-----------NSHEFISCRIFTLSFHYSLMANYCWILMEGLYLYNII
V. jacobsoni 3    FISGVGLYGTF-----------DDNQKSICCKLFLAMFHYTLMANYCWILMEGLYLHSLV
V. destructor     ---------ST-----------RHSDLTPQCWMLMCLFHYSLMANYCWILVEGLYLHNLI
V. destructor 2   FISGVGLYGTF-----------DDNQKSICCKLFLAMFHYTLMANYCWILMEGLYLHSLV
V. destructor 3   FTRGIT---SF-----------NSHEFISCRIFTLSFHYSLMANYCWILMEGLYLYNII
I. scapularis     FIDGVGLSTNV-----------DFNEENVDCKVFTSFWHYVLMANYCWILMEGLYLHSLV
T. urticae        F---------TL-----------EYNS-TLSCKLIIVWWQYSLLANYNWLLMEGLYLHNLV
T. castaneum      FIEGTALPHEIIQINGKLVY--NKTNFSWVCKAIISLWNYFIISNYMFLLMEGAYLHNLL
T. castaneum 2    FVEGLGLPSDLNYRNGSLYFDINSETNNWACKLLTSLWQYFITANYSWILMEGLYLHNLI
A. mellifera      FVSGIALASDVIIKNGKIYWLVDKKESNWLCKMFTSFWQYFILANYFWILMEGLYLHNLV

V. jacobsoni      FNSFYAD-GSSIAKYVVMGWGLPVSCVIVWALARNIVGNDTCWTSDAK--LIQWLLRGPI
V. jacobsoni 2    FMSVYAD-NSKIAKYIVMGWGLPLPWVAIWALARLCYDNNRCWEMEPKYRGITWILRGPI
V. jacobsoni 3    FHSLGND-PSSISKYTIMGWGLPVLFIAPWSLARALWENKLCWTTNKI-GWHEWIIRGPI
V. destructor     FNSFYAD-GSSIAKYVVMGWGLPVSCVIVWALARNIVGNDTCWTSDAK--LIQWLLRGPI
V. destructor 2   FHSLGND-PSSISKYTIMGWGLPVLFIAPWSLARALWENKLCWTTNKI-GWHEWIIRGPI
V. destructor 3   FMSVYAD-NSKIAKYIVMGWGLPLPWVAIWALARLCYDNNRCWEMEPKYRGITWILRGPI
I. scapularis     FLAFFTD-SSGILRYVALGWGLPVLFIISWVVARATLDDTLCWTTNVR-QDLFWIIRGPI
T. urticae        FFNIFND-NSSILKYIILGWSLPVLFIIPWIIARSLYEDTFCWVINKN-VALFWIIRGPI
T. castaneum      FLKLLSE--NGVVIYYSLGWGIPLLFIIPWIVLKAGNENIYCWTTKSS-KFIAMLIDVPI
T. castaneum 2    FRALFADSSNSIKWYVVMGWGLPLIIVGFWVAARLLVEDNLCWTTHEN-YDVFLIIGIPT
```

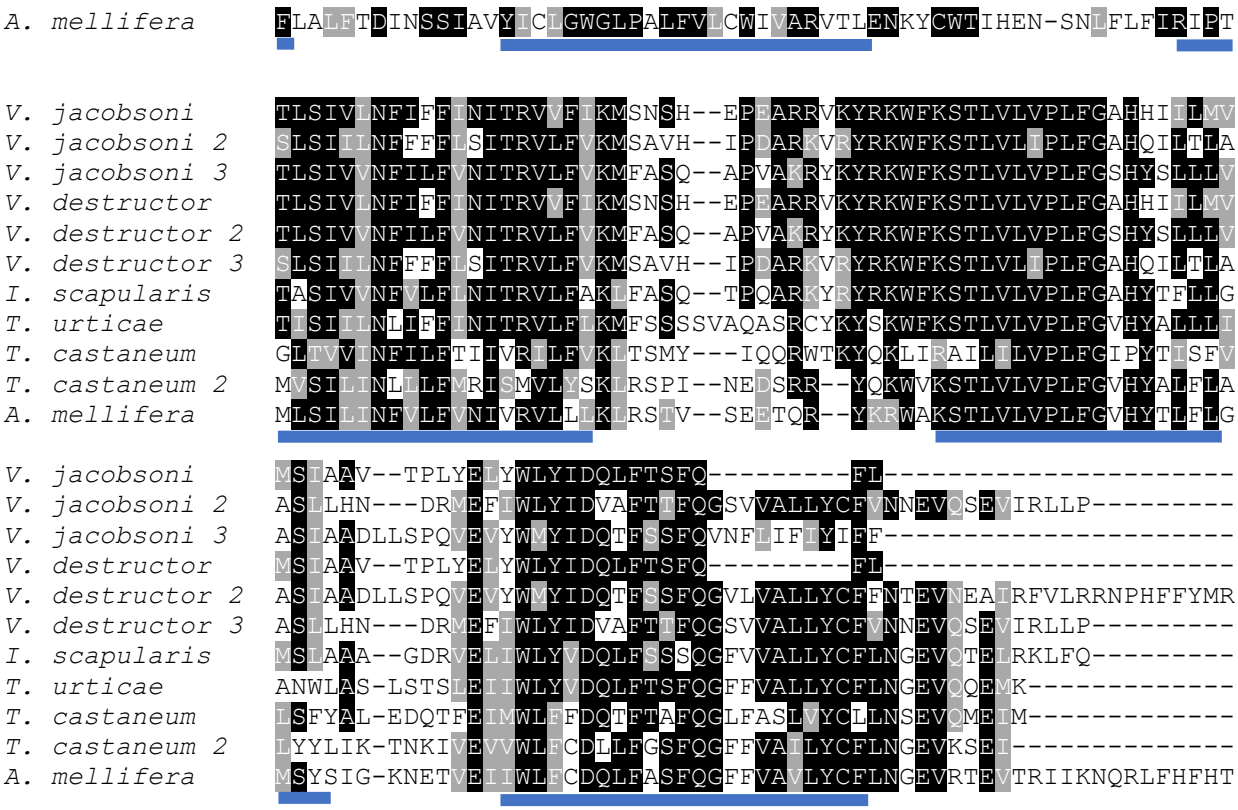


Figure 3.25 A) Gene structure for iPTH receptor sequence in V. *destructor* and V. *jacobsoni* is shown. The 7 Transmembrane helixes are indicated by blue color highlighting and underlining. B) Multiple sequence alignment of iPTH receptor from various orthologous species across Phylum Arthropoda.

Chapter 4 - Conclusion

The *Varroa* mite, an obligatory ectoparasite of honey bees, has had a significant impact on the decline of commercially managed honeybee populations in the United States and around the world over the last decade. While Asian honeybees are more tolerant of mite infestations due to their behavior and hygienic practices, European honeybees are severely affected by mite infestations (Haber et al., 2019). The mite-infested honey bee colonies displayed a variety of symptoms, including a smaller abdomen, deformed wings, and a shorter worker bee life span.

The current mite management strategy in commercially managed bee colonies is based on synthetic and natural acaricides, but it has proven ineffective due to the development of acaricidal resistance in these mites (Hillesheim et al., 1996; Maggi et al., 2009; Milani, 1995, 1999). It is difficult to develop/find novel anti-mite chemistries due to various limiting factors such as the direct impact of acaricides on honeybee health and the indirect impact on human health because many apiculture products are used for human consumption. There is an urgent need to develop novel mite control methods because the honeybee population in the United States is declining at an alarming rate; approximately 40% of commercially managed bee populations have declined since 2006. (source:www.beeinformedproject.org). New approaches have been sought that focus on solving the problem by utilizing new biotechnological resources and available *Varroa* mite and honey bee genomic sequences. Varroa mite-specific genes, for example, were targeted using RNA interference in conjunction with genetically engineered endosymbionts (Leonard et al., 2020).

Another possible approach is to identify and target the *Varroa* mite specific neuropeptide signaling system, which is absent in honey bees (Jiang et al., 2016). The rationale behind this approach is based on comparative genomics between honey bee, and *Varroa* mite. A

neuropeptide signaling system found in *Varroa* mites but not in honey bees' genome could be an ideal target (Jiang et al., 2016) for developing peptidomimetic-based acaricides (Altstein, 2004; Zhang et al., 2011).

Hitherto, *Varroa* mite neuropeptide signaling system is understudied area. Using comparative genomics, this study aimed to gain insight into the neuropeptide signaling system and identify suitable targets for peptidomimetic-based acaricide development. Using known neuropeptide and neuropeptide receptor gene sequences from *D. melanogaster,* a genome-based homology search was performed in the genomes of *V. destructor* and *V. jacobsoni*. Each of these species has 29 and 31 genes that encode for neuropeptides, respectively. Orcokinin and Myosuppressin are two additional genes that were found only in the genomic scaffold of *V. jacobsoni* but yet missing the computational annotation. The identified gene sequences shared an orthology relationship with their counterpart genes from closely related arthropod species. As our reference genes, known neuropeptide genes from *D. melanogaster, A. mellifera, I. scapularis*, and *T. urticae* were used. The orthology of the gene sequence could be described in several ways, including the location of mature peptides, the number of mature peptides, and conserved sequence motifs in the prepropeptide sequence. In the case of the neuropeptide receptor, we were able to identify 37 cognate GPCRs in each of *Varroa* mite species.

There were some anomalies in the receptor results. The receptor sequences for the neuropeptides sNPF, RYamide, Pyrokinin, and Gonadulin were identified, but not the neuropeptide sequence itself. In the case of FMRFamide, however, we were only able to identify the neuropeptide sequence and not the receptor sequence. During the evolution of the neuropeptide signaling system, both the neuropeptide and its cognate receptors co-evolve (Park et al., 2002). As a result, it is unlikely that a neuropeptide is present but its cognate receptor is

absent in a species, and vice versa. These anomalies are thought to be the result of an incomplete genome sequence or a missing part of the genome assembly. Even two neuropeptides, orcokinin and myosuppressin, were found in *V. jacobsoni*, while the cognate receptor for myosuppressin was found in both species while the cognate receptor for orcokinin is unknown (Caers et al., 2012). As a result, the study's findings raise concerns about the quality and completeness of the *Varroa* mite's genome sequence (Techer et al., 2019).

Comparative genomics between honeybee and *Varroa* mite

When the major components of neuropeptide signaling in *A. mellifera* and *Varroa* mite are compared, neuropeptide genes such as ACP, Allatostatin B, Arginine vasopressin like, Carausious like peptide, EFLamide, Glycoprotein hormone Beta 5 (GPB5), Natalisin, NPF, and Proctolin are found in *V. jacobsoni* and *V. destructor* (except GPB5) while they are lacking in *A. mellifera*. Through functional characterization and genetic studies, these targets can be studied in depth for their physiological and biological importance in *Varroa* mites. Appropriate peptidomimetic-based acaricides can be developed to disrupt the *Varroa* mite specific neuropeptide signaling system, resulting in altered physiological or behavioral outcomes and, ultimately, aiding in mite infestation reduction in honey bees.

Proctolin

Proctolin is a bioactive pentapeptide with a highly conserved sequence of RYLPT (Arg-Tyr-Leu-Pro-Thr). Although Proctolin has a strictly conserved sequence, there are few variations, such as RYLPA featuring T5 replaced with A5 (A = alanine), RYLST having P5 replaced with S5 (S= serine), and RYLMT having P replaced with M (M= methionine), in the case of *Oryctes borbonicus, Coccinellidae,* and *Daphnia pulex*, respectively (Dircksen et al., 2011; Spittaels et al., 1995; Veenstra, 2019a). It has the property of modulating interneuronal and

intermuscular synaptic transmission in a wide range of arthropods, such as the horseshoe crab (Groome et al., 1990). In insects, proctolin is associated with modulating contractions of both somatic and visceral muscle contractions (Orchard et al., 1989; Orchard et al., 2011). The mode of action of proctolin is believed to be through the elevation of both cyclic adenosine monophosphate, cAMP, and inositol triphosphate and increasing calcium entry through the voltage-gated sodium ion channels (Erxleben et al., 1995; Hiripi et al., 1979; Johnson et al., 2003; Lange, 1988; Lange et al., 1987).

Interestingly, proctolin signaling is missing from the insect orders Lepidoptera and Hymenoptera, including the honeybee, and is lacking in vertebrates including human. This neuropeptide signaling is found only in arthropods based on sequence homology, showing myotropic activity. Thus, an ideal acaricidal product can be designed using this neuropeptide signaling as its signaling is only limited to arthropods, so acaricides with no adverse effects on human health can be developed and used with fewer restrictions. An additional/separate study was conducted using the same principle in Park's laboratory. Several synthetic analogs of RYLPT neuropeptide were tested for their activity against the *V. destructor*'s proctolin receptor through a calcium ion mobilization-based luminescence assay. Further details are beyond the scope of this study and will be provided in the detailed publication soon.

ACP

The ACP neuropeptide signaling system was first identified as an AKH related peptide from *Locusta migratoria* in 1999 (Siegert, 1999) and then again in 2006 from *Anopheles gambiae* (Kaufmann and Brown, 2006), with unknown functions. This peptide considered to be the same group as AKH, was renamed as ACP for the first time in *A. gambiae* in 2010 (Hansen et al., 2010). The origin of ACP neuropeptide is believed to be the result of duplication of the

ancestral gene of AKH. Despite being an intermediate in structure between AKH and corazonin, ACP does not activate AKH and corazonin receptors and vice versa in *A. gambiae* (Hansen et al., 2010). Thus, AKH and ACP are considered as an independent and paralogous signaling neuropeptide system. We were able to identify a gene sequence that is very similar to the ACP receptor and was verified by black blasting and phylogenetic clustering. ACP neuropeptide signaling is also present in other arthropods such as *Tribolium castaneum*, *Aedes aegypti*, *Culex pipiens, Bombyx mori* but is absent in *D. melanogaster*, *A. mellifera*, *Daphnia pulex,* and *Pediculus humanus* (Hansen et al., 2010). Recent reports suggest that the ACP and AKH are involved in maintaining energy hemostasis by regulating the carbohydrate and lipid levels of the hemolymph in *G. bimaculatus* (Zhou et al., 2018).

Interestingly, we could not find the AKH gene sequence from the genomic scaffold of *V. destructor* and *V. jacobsoni*. In the absence of AKH signaling, there is the possibility that ACP signaling might be associated with and regulate various biological functions, including locomotion activities and energy homeostasis through carbohydrate and lipid metabolism regulation in the *Varroa* mite. As it is known, maintenance of the nutrient level in the blood is vital for animals' health; improper regulation leads to diseases such as diabetes and hyperlipidemia (Morton, 2006). Similarly, in arthropods such as insects, AKH and ACP are essential for the regulation of energy homeostasis. Ablation of the normal function of the neuropeptide signaling system affects energy homeostasis and food intake behavior. ACP neuropeptide signaling is absent in the honeybee and ACP neuropeptide does not interact with the AKH receptor and vice versa, which means very low activity against the homologous GnRHR of humans. Thus, we can develop human-safe peptidomimetic-based acaricides targeting the ACP signaling system in the *Varroa* mite.

Allatostatin B/MIP:

Allatostatin B (Ast B), also called Myoinhibitory peptide (Blackburn et al., 1995; Schoofs et al., 1991) or Prothoracicostatic hormone (Yamanaka et al., 2010), is known for various functions, such as inhibitors of JH biosynthesis, neuromodulatory activity, and myoinhibitory activity. Ast B signaling is widely distributed among insect species, including *Carausius morosus* (Lorenz et al., 2000), *D. melanogaster* (Williamson et al., 2001), *G. bimaculatus*, *L. migratoria* (Schoofs et al., 1991), *Manduca sexta* (Blackburn et al., 1995). Additionally, Ast B/MIP functions as a receptor for the Sex Peptide in *D. melanogaster* (Kim et al., 2010). Ast B has been isolated from a variety of other arthropods, including arachnids and crustaceans, for their property of inhibiting muscle activity (Schoofs et al., 1991; Lorenz et al., 2000; Davis et al., 2003; Kim et al., 2006). Similarly, Ast-B/MIP signaling is involved in regulating the salivary glands and hindgut movements (Šimo and Park, 2014). As Ast B neuropeptide signaling is absent in *A. mellifera*, it is a potential target for acaricidal development, provided upon understanding the physiological function associated with its signaling in the *Varroa* mite.

AVP-like:

The arginine vasopressin-like neuropeptide is a highly conserved metazoan nonapeptide that is known by a variety of names in different taxa, including oxytocin (OT) in mammals, annetocin in annelids, conopressin in molluscs, nematocin in nematodes, and inotocin in insects. Although, AVP-like neuropeptide signaling appears to be involved in neurotransmission, metabolism, and osmoregulation in molluscs and annelids (Gruber, 2014), the biological significance of AVP's signaling in insects remains unknown. Further functional characterization studies are necessary to ascertain the biological significance of AVP-like neuropeptide signaling

in the *Varroa* mite and to assess its suitability as a target for peptidomimetic-based acaricide development.

Carausious like peptide:

Caraursious like peptide is a novel neuropeptide isolated from *Carausious morosus,* a stick insect (Liessem et al., 2018). Due to its recent identification, we lack sufficient information about its physiological significance in arthropods. Further genetic studies are required to gain a more complete understanding of arthropod signaling.

EFLamide:

EFLamide (Glu-Phe-Leu-amide) is an arthropod neuropeptide ortholog of the vertebrate thyrotropic releasing hormone (TRH). Across species, the EFLamide gene is believed to encode two distinct types of peptides, namely EFLamide and EFLGGPamide (Glu-Phe-Leu-Gly-Gly-Pro-amide) (Veenstra et al., 2012). Among arthropods, the EFLamide gene was first identified from spider mite *T. urticae*, and was later discovered in other chelicerates such as ticks, *I. scapularis* and *Rhipicephalus spp.,* the mite *Sarcoptes scabiei,* the scorpion *Hadrurus gertschi,* and crustaceans such as *Daphnia pulex* and *Marsupeneus japonicus* (Veenstra et al., 2012) and in insects including *Locusta migratoria, Cimex lectularis,* and *Anax spp.* (Veenstra, 2019b; Veenstra and Šimo, 2020).

TRH is a member of the cysteine knot protein family found in vertebrates and is involved in a variety of functions, including growth, reproduction, development, and metabolism. Hitherito, the functional characterization of EFLamide in arthropods is yet to be done. It will be interesting to determine the neuropeptide's physiological significance in arthropods. However, if we target this neuropeptide signaling system, we will have to consider off-target effects, as EFLamide is vertebrate ortholog of TRH.

Neuropeptide F:

Neuropeptide F (NPF) was discovered in a tapeworm among the species in the Kingdom Animalia (excluding vertebrates)(Maule et al., 1991; Nässel and Wegener, 2011) and the first NPF receptor was identified in *D. melanogaster* (Garczynski et al., 2002). NPF shares a C-terminal sequence with vertebrate Neuropeptide Y peptide (NPY), with F at the last position, and a consensus C-terminal sequence with GRPRFamide (Gly-Arg-Pro-Arg-Phe-amide). The length of mature peptide of NPF varies significantly, ranging from 8 amino acids short to 34 amino acid long. In vertebrates, NPY signaling is associated with the stimulation of food uptake (Nässel and Wegener, 2011). Similarly, NPF signaling has been linked to modulation of feeding and reproduction in animals (excluding vertebrates) (Orchard and Lange, 2013). The presence and association of the NPF in buccal and cerebral ganglia in A. californica, as well as in neuronal cells related to egg-laying hormone in *H. aspersa,* suggests that the NPF may play a role as a feeding and reproductive modulator in protostomian species (López-Vera et al., 2008; Walker et al., 2009).

Misexpression or suppression of the expression of the NPF in neurons correlates with food seeking and aversion behavior in *D. melanogaster*. *D. melanogaster's* transgenic larvae with lost NPF signaling exhibit old larval behaviors such as food aversion and reduced mobility (Wu et al., 2003). Furthermore, the NPF titer level in the hemolymph of the female mosquito, *A. gambiae,* fluctuate before and after the blood meal, as well as during reproductive cycle (Garczynski et al., 2005). As a result of the biological functions associated with NPF signaling reported by various studies, makes NPF signaling is a potential target for peptidomimetic-based acaricide development, as chnages in feeding and reproductive behavior will have significant

impacts on the pest population. However, functional studies are required to confirm the role of NPF signaling in the *Varroa* mite.

Glycoprotein hormones:

Glycoprotein hormones are well conserved and characterized both structurally and functionally in the case of vertebrates. These hormones are members of the cysteine knot protein family, which includes follicle-stimulating hormone (FSH), luteinizing hormone (LH), thyroid stimulating hormone (TSH), and chorionic gonadotropin (CG). In vertebrates, these are associated with various functions such as growth, reproduction, development, and metabolism. The glycoprotein hormone occurs in the biologically active states as a heterodimer composed of two polypeptide chains. Following the completion of the human genome sequencing, two novel genes were discovered in mammals and named GPA2 and GPB5 based on the order of discovering (Hsu et al., 2002; Rocco and Paluzzi, 2016).

Glycoprotein hormones do have homologs in many invertebrates, specifically Bursicon and GPA2/GPB5. Bursicon is involved in a variety of physiological processes, including new cuticle formation, wing expansion (Honegger et al., 2008), cuticle sclerotization, and melanization. GPA2/GPB5 neuropeptide signaling evolved long before the emergence of bilateral metazoans (Hauser et al., 1997; Vibede et al., 1998), indicating that the conserved nature holds vital physiological roles in arthropods and other animals. GPA2/GPB5 neuropeptide signaling has been linked to the development and maintenance of ion and water balance in insects. The expression of mRNA transcript of LGR1 (receptor for GPA2/GPB5 heterodimer) was seen in various osmoregulatory tissues (Malpighian tubules, midgut, and hindgut) in *D. melanogaster* (Chintapalli et al., 2007; Sellami et al., 2011; Vandersmissen et al., 2014) and *Aedes aegypti* (Paluzzi et al., 2014). In the case of third instar larvae and pupal *D. melanogaster*,

knockdown of mRNA transcript of the receptor for GPA2/GPB5 led to mortality (Sellami et al., 2011), only a low number of flies could survive the transition from one stage to another. Even after the ablation of the endocrine neuronal cells expressing the GPA2 and GPB5 neuropeptides, the died because they could not survive the transition from one developmental stage to next. Another exciting fact is that GPA2/GPB5 neuropeptide signaling is sex-specific, i.e., males have higher GPA2/GPB5 signaling expression than females (Sellami et al., 2011).

The physiological role of GPA2/GPB5 neuropeptide signaling is a powerful target for acaricidal development. However, because GPA2 and GPB5 are arthropod orthologs of the vertebrate TSH signaling system, peptidomimetic acaricides targeting GPA2 and GPB5 signaling in *Varroa* mites may have unintended consequences for non-target species including humans.

Advantages and disadvantages of the approach

The development of acaricides based on peptidomimetics has several advantages. These benefits include improved metabolic stability, improved bioavailability, highly specificity in targeting specific pest functions, lower non-target effects, and lower production costs than conventional pesticide. In general, when compared to the broad-spectrum product applications, the pharmacological properties will be enhanced and optimized for efficient action against the specified target. Furthermore, peptidomimetic-based compounds can be made resistant to degradative enzymes by modifying minute but critical components such as a single amino acid mutation.

Second, comparative genomics involves comparison of different species at the genomic level and the provision of insightful information that can be used to understand the evolutionary changes between two different species. Recent reports indicates that the genome of *D.*

melanogaster is 60 % homologous to that of *Homo sapiens* (Mirzoyan et al., 2019). Thus, genes that are highly conserved across species can be identified, as can genes associated with unique physiological and behavioral responses in specific species. Following this approach, highly species-specific chemistries for protecting the beneficial insect or honey bee from parasitic infestation can be designed based on the simple principle of targeting the set of genes present in the parasite but absent in the non-target species/beneficial insect/arthropod.

Despite the many benefits of this approach, there are some drawbacks, such as the lack of annotated genomic information for various species. Although biotechnological advancements have aided in the acquisition of genomic information for various species, sufficient work in terms of genome sequencing for various arthropod pests remains to be done in order to find solutions to pest infestations.

References:

1. Altstein, M., 2004. Novel insect control agents based on neuropeptide antagonists. Journal of Molecular Neuroscience 22, 147-157.
2. Caers, J., Verlinden, H., Zels, S., Vandersmissen, H.P., Vuerinckx, K., Schoofs, L., 2012. More than two decades of research on insect neuropeptide GPCRs: an overview. Frontiers in endocrinology 3, 151.
3. Dircksen, H., Neupert, S., Predel, R., Verleyen, P., Huybrechts, J., Strauss, J., Hauser, F., Stafflinger, E., Schneider, M., Pauwels, K., 2011. Genomics, transcriptomics, and peptidomics of Daphnia pulex neuropeptides and protein hormones. Journal of proteome research 10, 4478-4504.
4. Erxleben, C., DeSantis, A., Rathmayer, W., 1995. Effects of proctolin on contractions, membrane resistance, and non-voltage-dependent sarcolemmal ion channels in crustacean muscle fibers. Journal of Neuroscience 15, 4356-4369.
5. Hansen, K.K., Stafflinger, E., Schneider, M., Hauser, F., Cazzamali, G., Williamson, M., Kollmann, M., Schachtner, J., Grimmelikhuijzen, C.J., 2010. Discovery of a novel insect neuropeptide signaling system closely related to the insect adipokinetic hormone and corazonin hormonal systems. Journal of Biological Chemistry 285, 10736-10747.
6. Hillesheim, E., Ritter, W., Bassand, D., 1996. First data on resistance mechanisms of *Varroa jacobsoni* (Oud.) against tau-fluvalinate. Experimental & applied acarology 20, 283-296.
7. Hiripi, L., Rozsa, K., Miller, T., 1979. The effect of proctolin on the adenylate and guanylate cyclases in the Locusta brain at various developmental stages. Experientia 35, 1287-1288.

8. Jiang, H., Kim, D., Dobesh, S., Evans, J.D., Nachman, R.J., Kaczmarek, K., Zabrocki, J., Park, Y., 2016. Ligand selectivity in tachykinin and natalisin neuropeptidergic systems of the honey bee parasitic mite *Varroa destructor*. Scientific reports 6, 1-8.
9. Johnson, E.C., Garczynski, S.F., Park, D., Crim, J.W., Nässel, D.R., Taghert, P.H., 2003. Identification and characterization of a G protein-coupled receptor for the neuropeptide proctolin in Drosophila melanogaster. Proceedings of the National Academy of Sciences 100, 6198-6203.
10. Lange, A.B., 1988. Inositol phospholipid hydrolysis may mediate the action of proctolin on insect visceral muscle. Archives of insect biochemistry and physiology 9, 201-209.
11. Lange, A.B., Orchard, I., Lam, W., 1987. Mode of action of proctolin on locust visceral muscle. Archives of insect biochemistry and physiology 5, 285-295.
12. Leonard, S.P., Powell, J.E., Perutka, J., Geng, P., Heckmann, L.C., Horak, R.D., Davies, B.W., Ellington, A.D., Barrick, J.E., Moran, N.A., 2020. Engineered symbionts activate honey bee immunity and limit pathogens. Science 367, 573-576.
13. Maggi, M.D., Ruffinengo, S.R., Damiani, N., Sardella, N.H., Eguaras, M.J., 2009. First detection of *Varroa destructor* resistance to coumaphos in Argentina. Experimental and Applied Acarology 47, 317-320.
14. Milani, N., 1995. The resistance of *Varroa jacobsoni* Oud to pyrethroids: a laboratory assay. Apidologie 26, 415-429.
15. Milani, N., 1999. The resistance of *Varroa jacobsoni* Oud. to acaricides. Apidologie 30, 229-234.
16. Mirzoyan, Z., Sollazzo, M., Allocca, M., Valenza, A.M., Grifoni, D., Bellosta, P., 2019. Drosophila melanogaster: a model organism to study cancer. Frontiers in genetics 10, 51.
17. Morton, G., 2006. Cummings DE, Baskin DG, Barsh GS, Schwartz MW. Central nervous system control of food intake and body weight. Nature 443, 289-295.
18. Siegert, K.J., 1999. Locust corpora cardiaca contain an inactive adipokinetic hormone. FEBS letters 447, 237-240.
19. Spittaels, K., Vankeerberghen, A., Torrekens, S., Devreese, B., Grauwels, L., Van Leuven, F., Hunt, D., Shabanowitz, J., Schoofs, L., Van Beeumen, J., 1995. Isolation of Ala1-proctolin, the first natural analogue of proctolin, from the brain of the Colorado potato beetle. Molecular and cellular endocrinology 110, 119-124.
20. Techer, M.A., Rane, R.V., Grau, M.L., Roberts, J.M., Sullivan, S.T., Liachko, I., Childers, A.K., Evans, J.D., Mikheyev, A.S., 2019. Genomic analyses of sibling honey bee ectoparasitic mite species show divergent strategies of adaptation. bioRxiv, 512988.
21. Veenstra, J.A., 2019. Coleoptera genome and transcriptome sequences reveal numerous differences in neuropeptide signaling between species. PeerJ 7, e7144.
22. Zhang, Q., Nachman, R.J., Kaczmarek, K., Zabrocki, J., Denlinger, D.L., 2011. Disruption of insect diapause using agonists and an antagonist of diapause hormone. Proceedings of the National Academy of Sciences 108, 16922-16926.

Appendix A - CNMamide

V. jacobsoni

Identified from TSA database : GETP01211115.1

Potential Prepropeptide:

MTGRSSAKRRTALCLLALLLFFRLSASSVVGEPIPSEYYGKTSDNDFILLNKLKELFEEKQYVA

EREKELDKERMKIQSIIMEGKESEIQSDSDYYAEKLPTPNAVVRQEPQHSGKRTMISYMTLCHF

KICNMGRKRQFK

Signal peptide = MTGRSSAKRRTALCLLALLLFFRLSASSVVG
Mature peptide = TMISYMTLCHFKICNM
Accessory protein =
EPIPSEYYGKTSDNDFILLNKLKELFEEKQYVAEREKELDKERMKIQSIIMEGKE
SEIQSDSDYYAEKLPTPNAVVRQEPQHSG, QFK

V. destructor

Identified from SRA database: SRR8864012.46893032.2

Potential Partial Prepropeptide:

ESEIQSDSDYYAEKLPTPNAVVRQEPQHSGKRTMISYMTLCHFKICNMGRKRQFHK

Mature peptide: TMISYMTLCHFKICNM-amide
Accessory protein: ESEIQSDSDYYAEKLPTPNAVVRQEPQHS,QFHK

'No match in Refseq Genomic Database so could not predict gene structure

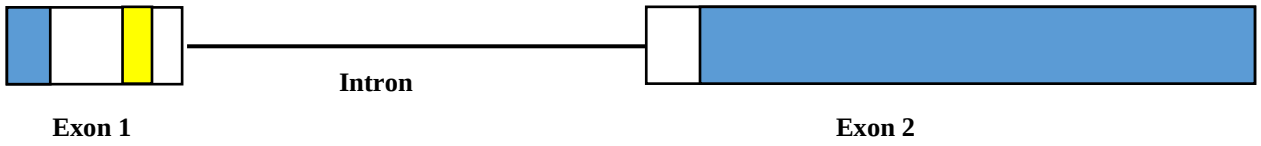
1. .

Appendix B - Corazonin

V. jacobsoni

Identified from: Refseq genomic database and TSA database

Gene structure:




```
TGAATCGGTTTTTGAATAATTATGTCTCATTTACATACATCGGGGCGCTAGTAGAAAATCCCAGCGACGTCACTGGC
GGATAGTTTGTCTAGCCTAAACTGGAGTTGGTGTTCTTCTGTTTGAAGTGCGGGCGGTTCACCCTCTCCCAAGGGTG
ATATATATAAAGGGGTGTATAAAAGCTCAAGATCCAGTCGACATACATTTAGTGTGCTCCTGATCACCTTACAGAGT
GTGCCGGCTTGTAGTCGCTTCGAATCCAACAATGAAATCCGGTCTTGCAATTCTACTCGTGTTGGCAGCCCTAGTGG
                              M  K  S  G  L  A  I  L  L  V  L  A  A  L  V
TTATGTGTACCGCGCAGACCTTTCAGTACAGTCGGGGCTGGACCAATGGCAAGAGGGCTATCGGTCCGGTTGATGTC
V  M  C  T  A  Q  T  F  Q  Y  S  R  G  W  T  N  G  K  R  A  I  G  P  V  D  V
                ⇓
GACGGCAAGCTGGATTCGGGACGGATAACCGCCAAGGATCTCTTTACTCTGGTCGAGTTGAATCGTCGAATGTGTCT
 D  G  K  L  D  S  G  R  I  T  A  K  D  L  F  T  L  V  E  L  N  R  R  M  C  L
ATTCCTTGCCGGTGGTTCGCACGAAGACCAGTGATCGCACGGAAGCCAAAACTCCTAGAAGATAACGGATCAGTCAG
  F  L  A  G  G  S  H  E  D  Q  *
GACTACATCGCACCTGGTTGATCATTTTCTATCAATGATTATCGATAAATCAGTGAAGTTTCAGGTAAAGGGATGTT
CTATCACGTTTGTTACCGCCAGCGCTTTCGCATGTGCGTACTAGATAGTCGTTTGGGGGAGGAGGGGTGGAAAATCT
AGCGAACGCTACTGACTTTGAAATAGGTGGGAAGCGATCAAGAATGACAATGCGTTGCCTAAGTATTACTCCTGCCC
ATTGCAGTAGCTGGCTGATTAGCTGAGAATAGTTAGTCTTGTTTATTTCATAGGAACCAAGTAAGGCGGAAATAATT
TAGTTTAGCTTGAAAATATCTTGTTTCAATGAATCGACATTGTAGGAAAAAAGTAAAACACCTAGGGAAAGATTACA
CATCACCCAAAACACGACATTACCATCCAAGAATGAGTAAACGTCAGTAAAATGGAAAGGCGGACAGAAGACGATAA
ATTGAGAAAAACAGAAGCGGCACCCTAGGACTGCGGAGATAACCCATTAGTGGAAGCGTAATTTATCATACAAGTAT
TGGGTGAGCTGTCAGCAACAAACGAACACGACCATAGTGTGAGTACGAGTCAAAATGTATCAGTAAACCCATGGCAG
TGGAATATGTAAAACTTCAGTGTTCAGTCGGCTCAATAAGGCAGCTCTGATGGCGGTCAGAAAACCACAGGAAACGC
CGACAAACGTCATGAGAATACCTCTAGATGGGGACTAATTTGGCTGCTATCGATCATTGCACTAACCTTCAATCATG
CACCCAATCAACAGATTAAGCCAACAGAAAAGTGAACCACGTCGATCACGTTCCTCATGTAACGACGCCAGACATGG
GAAAGAGCCGGCCACGCGCTTGATTTATTTTCCGCACAGTTAATAAATGTGTTAATAAACAAATAATAAAATTACAA
ACGAAATACGCTGAAGCAAACGTTTTATGTTGTATTTTGTCTTAGACAATTTAGATTAGAACAGCCCAACAGTTGAG
GATAAGCTCTTCTTTGTCAGATTCTTGCTTTGATAAAAACCGCTAGGAATGTACGCCCCAACAACTTAAACACCCGA
GCATTTGATATAATTGAAAAATGTAAATAAAAAAAAAATCTGCGATAACCTTTCAGCGACAGAGAACTAGCTTCCTG
CTTTAAATTCATTATTATCTATTGAGCAATAGACGCTTTAATGATTCTGCACAACAATTATTCTTACTAATGCTAAG
CGTTGCAAAATAATCGAATATCTACAGTTTAATTAAGCTTGGCAAGAAATACGTTTACAAAAACTTGAGAATAGGTG
GCAGTTGTGCAAGTCAACGTACGTTAACAAAGTTTCCAGAGTAGTCGAACTATATATATTTCTGTATATATATTTAG
TCGCATTTATGAAGCACCGTTAACAGTTTTTTTTTATGTGAGACTTCTGCTTAGGACTCCTTTTTACGAAGTAAGGA
GTGCAGAGAAGAAAGGGGGAACTGTTAACTATTGTTGAGAACGGTTAATTCCGCTGCTTTCACATGATCGGAGCATA
TATAGAGTTGAAACTTTACAAAGGTTTACAAAGTCCTTTACAAAGGCACAACAACAATTACAAAGTTGTGTATAGCG
TGTAAACGAGAAATGAAGAAGCGAACGGAGGCTTATTATTCGATGAACATGACCAAACACAAAAAAATTTAAATAGC
ATTGCACATAATAATGTTATTACGTCATTAGGCTGTGTTACCAAAGAAGCTCTTCATTATCTTAAGATCCTAGCTAT
GGCTCCCTCTGTTACCTACGCTTAATACGTTCTTCAACCGTGGTAATAATCTCTTCATAATGCAGATGCACCTAACA
TTGCAAGAAAGTGGAACATTGTATATGTCTCTTTGTCGTCTCTGTGGCATCAGGGATGCCGAGCTATTATTTTTATT
AACGAATTCCTTAGATAGAGAGGAGGGTGGAACTATATGGGCTAAAATGTGTATAATGCAATAGTTGAAAACAATCA
GAGGCACGAGAGAGGAAAAAACTATATTCGCCAAAAACTCTTCTACTATGTAGCCTAATTATCTCACTACAATTTGT
TTATTTTTGAGAAATATGAATTGAAAACATTCATGTCGTTAACCCATATAGTTCCGTAAGGTAAAGTTATACGGACC
```

TGAGGTGGAACTAGATGGGTTATTCATGATTCCGTGCGAAGAAGTCACTCAGTCAACAAAATCAAAGAGAAATGTTT
TTCTCCTCATAAAGGGCTGTGATTATAACTAGAAAAAATGAAACTTTATAATCATTTACTCCAACCTCGCCAGTCAT
TCATTCACTGTAGGGCCATATATGGTGCGGCTGCGCCGGTGAAGCTGGATTTTGAATGTATAACTATTGGCACATTA
ACGCATTGTGTTTTGTTTAAGAAGTCGGAAGGAACAGTAACAATGTGCACTCCGTTCTCCGAAGCCAACTTCAGTGC
ATTAAAATCTGTATAATTTTCATGGATATGCATGATTTTAATATGAATAATTATCCATGATGAATCCTGATTTATGA
TCATGATTATCCACAATATGGTGCGGGTGTCGGTCATATTGTTTTGTATGAACGATGCGTTATGACCATACAATCGT
AGAAGATTTCTTTAAAAATAACCGCAACATCGACATGCGGTGCAACAAGCGCGAAAATACAATACCCGAGCTTAAGT
TTCGCTGTAAAGCTATTTAGTCCAATTTTTCTGGAAAATTAGCATCTACAACATATGTCTTCATATATGCTCGAATT
CCAATAGCCCATACACATCCCGTAGTAATTATGTAAGATAACTCTATATATGGTAAATTTTGTGAAACTTGTCTTCA
GAAATTTAGCTCTGGGATAGGAACTGTGGTAGAAAAACGGGCTGGTTATCTGTCTTTGTTAGGCTGTTACAAGCGAA
TCCTGTGATTTTTGAAGTTGAATAATTTTTAAATCATGCACCGTTAACATGTAAAATTTGGCTAGTAGTTACTGCAG
TCATTTGGAAACAGGTGAGTGAAAATCGGTTGAATTCGCTCCATTATAAAGGAGTTCATAAAAAAAAACATCCTTTA
TTCTCGAAAATCAATACTAAGTGATCAAGTTTTAGCCACTCGAAGGCCGCCATATTCACGCTGGAAATTCATAAAAT
ATTAGTGATGACGTTAGGCAAAAAGGCCCTAAGCAAAGAATGGCACACTAACTGAAGTTGAATCATGAAAGACTAGC
GAGGAGGTGACCAAATACTAGTTCCTGATGGACAGAAACGCATTGCCGTAATTAAAGAGGCTTTTTAGGAAAACAGG
ACTTAGAGCAAAAACTGGCATTTCAGGTGTAGTATGTTACAGAATTGTATTGAAAATACGGAAACTGAGTAAAGATG
CTATCAGCTGACATTCGACTCCAGGTCAGTTGGAAACGTATATGGTGTGCTTCATGCGCAATTTACGGAACAAAAAG
TCTACTTAAAAGCATATGATCGCCATAACCGAGACATGGTTCAACCACTACCACAAGCTGAAAAGCGATAAGGCGAT
TATTCTACTCTTAGCAGGCGCGATAGGCGTTGCGAAAACACACATGCGTCCAGTTCTAGCTAGCTTATGAGACTTGA
CGCCTACCTTATTTCACAAATCGTTCTCCCGGAAGAGTGAATTCCTAGTCCGAGCCGTTTAACAGTAGAGATTTTGG
ATTAGAAAATGCTTCATATAAAAATATGTAGAATTTTTGTGCTGTTTACATGTAAACAATAATAGTATGAATCACCT
AAAAAAAGCTTTAAGGAAAGAGTAGTCCAAGGCTTATTTAAGGCCTTCGGGAGAGTCCAGGAAGGAACGAATGCAAA
GGCTGAGTTTGTTCATACGGCTAGGTTGAAGAAGTGTTTTAAAGCAGCGGACAAATGATGGTGCACGCACATTGAAC
CTGGATATGATCAATTTGAAAAAGGCAACTACACCCCGGATGACTGAGACGTACGTAGGTGGTACGGAAGCAGATGT
GACAGTTTTTTAGTTACAGCTGCACGTAAAAAAAAAGAGTATTTTATTGTGGTTTAAGATTATATTTGTGATAATTT
TGATAATTCTGGTAATAGTTTACCGTTTATCTTCTTTATGGAACGAAGAAGATGCTCGTTGCATGCAACACAAACTG
CAAATAAACAAACGCTTAGTAGAATGAATGCATTAGAACGTTGATTCTGATAGTTTGCGTAGCACGTGTTATTTTGC
TGTAAAATATTTGGATACAGGTCTGCAACTAATGTGCCATGGTACAAATAAAGTTTGTTGTATATTTAAAGAAAACC
AATACATTTTGAATCTTCATGTTAGATCTTGGTATAGCTTATCAGCTTCTATGCTTTTTTATCAATAAAATACATCC
TCTAGTTTGTCATTGAATTTCCAAGAATAGTTACAATGAATTAATGAATCATCTAAATACGCAGTCTGTTTTAGAAG
TTTCGTTAGTCCCAGAAAGACGAATCGTGTTAGCAGCTGGAATTGGACCGACATCGCCTCGTGTTCTCTTCCAGCAG
GCAGACAAATTACCATACGTATAGACCCTAATATGATTGCTTACTCTTTGTGTCAGCTCAAATTTCATATTTTTAAA
TCTTCTTTTTAATAAACCTTCTTTTTGTAATCCTATAGGAGAACAGGATAACTATCGACTACTTCTAAAGCCAAATT
TCCGCCTATCACACAGACAACTGGCCTAGCTTCGGAGGACTGCTTAAACCTACTCTAATTTTGCAACTCGCTGAAGG
ATCTCTTTGAAAAATAAGACAGCCATGAAAATTAGTCGAAGTTATTTTGTATTCAGACAGATAGGTAATATACGGAT
CGAACAGGCTTTAAATGCAACATACATCTAGTCCAGCTCCACGTGGCAGTTTGATCTCAAGGTAAGTCTGACCCATG
ACTTTTGGGATACGCGTCACCTCACAGTACTTCACAGGTCGTCATCAAATTTTGAGTATTCGTAATAATTATTTAGT
TTGCTAAAGTTACGTTAAAAACCAAGTTTTACTTTATATGGATAAACTTGGGTAGACGGTTCAACAGCGACTCCTAT
TAGAAAACAAAGTTAATTGCTTACCTAAGCATCCGTGGATATTATAAACTATAAAGCGAGAGTAATAAAATATTAGC
GACTGAAGATCGCTAATGAATTTGACAAAAGTAAATTTGTGCTATCGCCCTACAATAGGTTTTGCGCGATTTGCAGA
CGTAGGTTTCACTTGACTTGCTGCAATTTCATGCACATTGATTCATAATCACGATTTCGTTACAGATTTAAGCTACG
GTAGACACCAAAAAATCGTACTCATAAAAAAATTACTCATACTTTCCGTAGTTTGCCGTCATTGGGTCAGTAGGGTA
TAATACAAACTGGCTTATTTGCGTTATCGATGCCCACAGATGTATGTTTACACAGATTAAACAAATCTCATCATTGA
AACTTCTACAGATCATTAGTTCATCCAATCTTGCGTGACAATTATCCGTTACATTTACCATAAAGAAACATAGCCCT
TCAGTAAATACTTTGCTACAATGAAGCCACGAAAGACACTTTTGTTTTGCGCCACGCCAAAACCTAGACACGCCTCC
TGTATCTGCGTGTTTCACTTCCGAGCGTTTAGTCGGGGCTCCGTCGATTAAGAGTCTTCCTCGTTCGACGTCGCTTC
ATTAGGAGGAAACTTTCCTCAATTTACAGGTGCTCTTCGGACACAGGTGACTAACTGAGAGTCTCTTGAGACTAATC
GCATCAGTTTGTAAAGGGTGTGCAATGCGTTAATACAGCGTCTTTGGCTGCGAACTCAGGCAAGGGAACAAAAAAGG
AGAGAGCCTTTTTACAGAGTGAGGCACTCGTTCTCTGACTGAGTACAATTGTTGCGCGGATTCGATGCTCTTCGAAA
ATAGACGCTTCTACATTGCGCTCTCATAATATCACAGTTACAGTATAGTTGTTATTCTGTTGCGGTTTGTCACTCGA
AGCCGCATAAACTACACTATTATTATGATTATAAACGATACGTTACTAGTATGACTGAAAAACTAAGTAATCAGAAA
GCTTATAAATTTCGACTGTTTACTAACGGCAAAATTAATTCGCTCTCTGTGAACTGTTGGCACAATTACCGGCGTTG
CGAGGTTGGTTTATGGTAATAGTAAGACGCTTGTTGATGTCTTATACCAG

Potential Preprohormone Sequence:

MKSGLAILLVLAALVVMCTAQTFQYSRGWTNGKRAIGPVDVDGKLDSGRITAKDLFTLVELNRR

MCLFLAGGSHEDQ

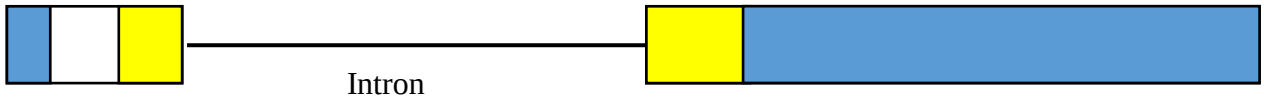
Signal peptide: MKSGLAILLVLAALVVMCTA
Accessory proteins: AIGPVDVDGKLDSGRITAKDLFTLVELN and MCLFLAGGSHEDQ
Mature peptide: QTFQYSRGWTN-amide

Appendix C - Diuretic Hormone 31

V. jacobsoni

Identified from genomic database and TSA database

Gene structure:



Blue region = Untranslated region
Yellow region = Mature peptide region

GGTCTGACAGTCTTCTTTCCAGCCGTTGTGGACTTCGTCGAG**ATG**ACTCAGATTCAAGTAGTTGTCGTTGTATTTCT
M T Q I Q V V V V V F L
TTTGGCTTTTGCCGGATGTTTCACTTCACCAGCACCGCGTAGCGACGAGGCTCTTCAGTACTATTACTTCATGCAGC
L A F A G C F T S P A P R S D E A L Q Y Y Y F M Q

ATCCTCCCAGCGTGGAGTTGATGCTTGGCGATAAGCGGTCGAACGGAATGTTGGACTTCGGGCTGGCGCGCGGCATG
H P P S V E L M L G D K R S N G M L D F G L A R G M
TCGGGCGTGGACGCTGCCAAGGCGCGTCTCGGTCTCAAGTACGCCAACGATCCCTATGGCCCCGGTCGGAGG**TAG**
S G V D A A K A R L G L K Y A N D P Y G P G R R *

ATCTTTGATTAGAAGCCAATTGCTTGCAGCAGGCAACACCCCACCTTCCCACGGACCTGAATAATGTACTACACGTT
TATTAATTATCCTGTACCCATTACGATTATTAGTATCACCTCTGTTAACTAGTAAAAACCTCCTTGTTAACGGGCTA
TGGCCTGACGCCATCTGGTGCACGCTTCCAAGAAAAGCACACGATCGTTATGAAAGACGCTAGTGCACCGAAAGCTG
ACGTGAGTCAGGCGTTTGGCCATCGTCCTACATTTTGGTCCCGAGAAAGAATGCACGTAGAAATTGTTAAAGGAACA
AACGAAATCGGGCGACGGTTGCGGAGGATTTGGTCAACGCCCAAGTAGATTAAACTGTCGACAAACAAGAGACACAA
AGTCGAAACAAATCAATACTCACCCCACAAGTAGAGAACTTTTACGATAACGATTGTTAGCACGTCGTTGGAACTTG
AAATGGGTAGGTGGGGTGGTAATCAGTTAAGCTAAACAATAGTAAAGAGGCTGGCTGATACACAGGCAGAATCGTGT
ACAGACACGGGTACAGGCGCGGTTGTGAAGACATCAGCTGTTGGAGAAGAGACATGCAGATTAAAGGACAAATACTG
GCTGTTGAAGAACACTACAAGAAACATATATAGACTGAGGAGCAAAGATCCAAGGACCTCCGATAATGTTAATGAAC
AACATCCACAAGAAAGGATTACATATCGATCAGCGAGACTTTACGGCATGTTGTCATAGCGACCCGATAACATATGC
TCATCTACGTTAGGTAACAGCAATATAATTTCAAGATGCGCTGACAGCACCCGAAACCTCAGATAGACCTTTAGTGA
CCTGCCCACACCAGAATGTAACCCCTGAAGTCATATATCGATCTATTGTACTTGTCCCGCACAGGATCTGGTTACAA
TCTACTGTATTACGATACTATTTTATCAGATAGCTACCGGCGCATCAGGCAAAATGCTACCGACCAATTTGCTAATG
ATTGGTATTTATAAACAAGATTCTCTAGCATTCACTAGAAGTAAGAGACCGCAGATCGAATGCAGTTTCTAGGGGAT
AGGTGGACACTTGCGCAGTGAGAGACCTTGCTCACCAGCAGTAATACTATTGATATCTCCAGATAATAGGTATACAG
ACTGTCGAAAATCGCGAGTCCAAGAGGCATATCCCTGTGAAAACTTTTTTGTCGCTGACGAACATAGCTGATGTGTT
GGAATTAGCTGAAGAATTGATGAATGTGTAGCCGTCATAAGCGAGCCAAACACAACCAGCCGACCAATAGATAGACA
ACAAATATCTACAACGCAACATGTTCGGTGCAGATAGTTAACAGATAAGTACAGAAAGAAACCAAACGAAAATGATG
TCATCAGCGGCTGTTTATGACGATACACGCAAAAAAGAAGAACAAACGAAATGATGTGGGAAAATGGAAGGGATGTC
ATTTGCTGCGTTTCGTTAGACGGAGCATGGTAGGGATGTCAAACAATACACGCTCCACAAAGGCTATAAAAGCAGCA
AAAAGACAGGAAGCTAAGGATTTGCTTCAGAACACTACACGATGGAAAATCAGAAACAGTTGACATAACCAAAACTG
CAACGACGGGAAAAGATAGAAAGATAACAAAGAGATTGTGTAATTTATATAAATAGCAACCAACATTTACTATAGAA
CGATACAAGCGAAACAAAAAAAAATAGCAGACGAAAAAATCCGAGCAAAAAACAAAAAAACCTATTGTTATTACTGC
AAGCACCATTATAAACTGTACGACATTAGCTGAATTGAATGAATGAACTCGAGTTCACTCACGTGTTTCAACTTGTC
ATGAGGTAGCAATCAATCTGTTATGGTATCATCAGCAATGACAGAACTAAATCTAGGGAGACTAACTGACATTAAAA

CTACAGGCCAACTAGTAAAAAGAGCTGAACGAGCGAATCAAGCAATGGTATCGGAGAGCAAAAGAACGCGATATCAC
TTCGTAGCTGGGGACTCCTAGAGGAACATCAATAGTTTTCACGAAGAGGAACGACTTGCAACGGACAACTCAAGCGA
AAATGCGGTCGGTCGCATAACGTCGAACAAGCGAATACGTTTCAGCATGAAGACAAATATTTACAAAAATATACGGT
GAATCAAACCCAAATTCGGCAAAAATAAAAGACGATACTTGTTTGCAGATATGGGCAAGCCCTTTAACTTTGTTATA
GACAATTAGTTCGACATAATGGTATAGGATAGTTACAAATACGAAGGCTAAGTAGCAGAATACTACACACACATATT
TACTGGACAGATGCTTCGTGACTTGTTTCAGATTGACCACTGACATTTCTCGCACTGGGCCGAAGACGAGGCTATGC
TTTTATAGGCACGATGATTGCCGCTCCATTCTCCCGAATCAACGATTTACCATTGTTATAAAGGCACAACATGTAGG
TGCTTCGTTTTACACAACAGCAAGCAAACCAATAATAACACAAGTACAGGATCATGGAAAACCTACTTCTAAAAACA
ACTGCACAAAAATAGGCCTTTCGACCGTAAACAGATACACTTAATAAGTACATTAGATTTAATAAACGTCGAGGCTG
CGTAACTGACGTAGATCGCTTCGTCTGAGACTTCGGCTAACCGGCAGCGCCGCATTGCGGCAGAATGGTATACTAGC
ATTCTAACCGCAGCGTAAACAACGTTCGCTCAGGTAACTACGGTCCAGTAGGTAAACGAATCGTATAAATCGCCGTT
GAAATTGGATCTTTAAGGTTCATTATCGATGTCCAATTAGAGCCTGTTGACTCTGCGGTGTCTTCGGCTATGACGAG
GCATTACTAAGGACGACGTATTCCCTCCCCCCCGCCCGCCCCTACGGAATACAGCCTCTAAGGTCGACAACACGCCT
TTGTACGGGTCGATCAAGCGAATCAAAGCCGTTTTCCAAACAATGAAAATCACAGCATCGATACGTCAAATCACAGC
ATTCTTCACGAACAATTCTGGTTCCACGTGCGTACACGCCCGTACACGAACGCTTGAAGCACTAAAATGACAGAGCA
AATCGATAAGGCAGCTTTGGGCTTTAACATAGTTAGCACGTATTATCAGCTTAACGGTTCTTGCATCTAGCCTTCTT
CTTCCATATAAAGAGACACGCACGCACCCTCTAGCTTAGCGTGAAATGAAGAGGTTTTCTGTGAATATCAAGTCGCT
GATCCATAAAAGTTGAAGTAAACAACGTTAGGCCACATACTTGTATGTATGGGGATGGAACAAGTGTAGCCTTATGG
TTTTTGGGTGGAGTAAATTATGAAGTAGGCCGTAAAAGGTCGCCTTACCCTTAGGAAAAATTAATTTTCAGTTATTT
AGAGGAAAAAATTGTGCGCTGTTAGCTATTGCAAGGATATTTTCAATATTCTAAAATGCAATGACTAAACTGTGAGA
GTGGCCGGTTCGAGTCCCGTCAATGAGACGGGAAGTGTAATTGAAGGGAGAAACGAACCGTGATCTGAAGTTCGCGA
CATTGCGTTGACTATTAACAGCCGGACAGACTGTGATGGAAGGCTCGTAATTATTCTATGGTTTACAGTTGTTATTT
AGCTGTCAACACGATTCGCTTCATTGAACTTTCATTTGATAGCCAGTAGCGGTTTAGAACTTGCGGCAACTTGACGG
CTTCAACCGTTACAATATGATTGATAGTAACGTGACATAAACGGGCAGTAATTTAACTTTAAAGATAGAATGTCTAG
AGTTTTAGGCACATCAGTAATGGTGCGTTCGTGAGCTTATTATTTACCTAGTAGTGTTACGCATGTACAGTATTAAG
CGCAGAGTAAATGTTGTATCTGTCTGTTAGCTTTGATGTTAGTCTGTATTTGGTAGTAACAAGTTTATTATAACAAG
AAACTCTATATTATTCAAATCGTTTGCTAAGAATGAGACGCCATTATTCTACTATATTTAGCAGATATGACTAACTT
ATCTATTGCTCGTTTGACACTTTGGCTGCAACCGCAACACTAAATTTTTCGAAATTGTCTCGAAATTTTCGATTTTA
TGCTACGACATACGAACCCTAGTTATCAGCATTTTCCAAAAGGCATGAAAAAA

Potential Preprohormone sequence:

MTQIQVVVVVFLLAFAGCFTSPAPRSDEALQYYYFMQHPPSVELMLGDKRSNGMLDFGLARGMS
GVDAAKARLGLKYANDPYGPGRR

Signal peptide = MTQIQVVVVVFLLAFAGCFT
Mature peptide = SNGMLDFGLARGMSGVDAAKARLGLKYANDPYGP-amide
Accessory protein = SPAPRSDEALQYYYFMQHPPSVELMLGD

Appendix D - Elevenin

V. jacobsoni

Identified from: TSA database, GETP01191846.1

No match in Refseq genome database so its difficult to predict gene structure

Potential preprohormone sequence:

MIIVAILLCYFHAQSESVNCELYPFHHTCRGTMSRKRAMFPIAYGSECEESKGNINCVKEFEEK
NRIAYIPLSKSKLLIALLDDDLRKDITRSIRHKLRNDEMIKQNSALMENFLSQLDSSDNY

Signal peptide: MIIVAILLCYFHAQSES
Mature peptide: VNCELYPFHHTCRGTMS
Accessory proteins: AMFPIAYGSECEESKGNINCVKEFEEKNRIAYIPLSKSKLLIALLDDDLR
KDITRSIRHKLRNDEMIKQNSALMENFLSQLDSSDNY

V. destructor

Identified from: SRA database, SRR8864012.149089260.1

SRR8864012.149089260.1
CGTGAATTGCGAATTGTATCCGTTCCATCACACGTGCAGAGGTACAATGTCGAGGAAACGCGCG
ATGTTCCCAATTGCGTACGGATCAGAATGTGAGGAA

Partial preprohormone sequence:
VNCELYPFHHTCRGTMSRKRAMFPIAYGSECEE

Mature peptide: VNCELYPFHHTCRGTMS

Appendix E - Limostatin

V. destructor

Identified from: SRA database, SRR8864012.120933285.2

No match in Refseq genome database, So no prediction of gene structure

Potential Preprohormone Sequence:

MWSTLAILCLITTFGFNNAERSLNSNSFSQLSDPTPSRLRNLEDVPARLSRIFGKNSRTRSLHI
EDQHGFDGKVEKINVNSEEDLPGIRAYGVQKKKLMNNGYIEKSIQREIDTRSVEDGSRVTRSIE
AYGKTKDRLETNGSIERGNPSTIHETRRSPRSIETETTKQDGVKESDLEDLEGQDAKVFRPLFV
YRQQMARKQHRAKNQIYGFPKIPCEKRAPIF-

Signal peptide: MWSTLAILCLITTFG
Mature peptide: SPRSIETETTKQDGVKESDLEDLEGQDAKVFRPLFVYRQQMA
Accessory proteins: FNNAERSLNSNSFSQLSDPTPSRLRNLEDVPARLSRIFGKNSRTRSLHIED
QHGFDGKVEKINVNSEEDLPGIRAYGVQKKKLMNNGYIEKSIQREIDTRSVEDGSRVTRSIEAY
GKTKDRLETNGSIERGNPSTIHET and QHRAKNQIYGFPKIPCEKRAPIF

V. jacobsoni

Identified from: TSA database, GETP01243268.1

No match in Refseq genome database, So no prediction of gene structure

Potential Preprohormone Sequence:

MWSTLAILCLITTFGFNNAERSLNSNSFSQLSDPTPSRLRNLEDVPARLSRIFGKNSRTRSLHI
EDQHGFDGKVEKINVNSEEDLPGIRAYGVQKKKLMNNGYIEKSIQREIDTRSVEDGSRVTRSIE
AYGKTKDRLETNGSIERGNPSTIHETRRSPRSIETETTKQDGVKESDLEDLEGQDAKVFRPLFV
YRQQMARKQHRAKNQIYGFPKIPCEKRAPIF-

Signal peptide: MWSTLAILCLITTFG
Mature peptide: SPRSIETETTKQDGVKESDLEDLEGQDAKVFRPLFVYRQQMA
Accessory proteins: FNNAERSLNSNSFSQLSDPTPSRLRNLEDVPARLSRIFGKNSRTRSLHIED
QHGFDGKVEKINVNSEEDLPGIRAYGVQKKKLMNNGYIEKSIQREIDTRSVEDGSRVTRSIEAY
GKTKDRLETNGSIERGNPSTIHET and QHRAKNQIYGFPKIPCEKRAPIF

Appendix F – Myosuppressin

V. jacobsoni

Identified from: SRA database, SRR3635105.57299974.2

No match in Refseq genome database, So no prediction of gene structure

Potential partial preprohormone sequence:
VIKKLYFIGFHESIPLLDSGVKRQDVDHVFLRFGRRR-

Mature peptide: QDVDHVFLRF-amide
Accessory proteins: VIKKLYFIGFHESIPLLDSGV

Appendix G – Neuropeptide F

V. destructor

Identified from: SRA database,

Gene structure:


```
TGTTTCTGTTACTCATTTTATAATAACAGCTAACCATGTCAGCCGAGATGTCAAAGACGATGTG
                                                 M  S  K  T  M  C
CCTTCTCCTGGTTGTTTTTGTAACGGTGATGGTCATGGGCGTGGCAATGGCTCAGGAAGCACCG
  L  L  L  V  V  F  V  T  V  M  V  M  G  V  A  M  A  Q  E  A  P
AATAATCTCAACGCGGCCCTCGAGTATCTTGAAGGTCTCGACCGTTACTATTCAGCGAAGGCTC
 N  N  L  N  A  A  L  E  Y  L  E  G  L  D  R  Y  Y  S  A  K  A
GACCCAGGTATGGCCGTAGCGTGGGTTCCGGGACCCAGCGTCTCTACACTGGATGCGTCAATCC
R  P  R  Y  G  R  S  V  G  S  G  T  Q  R  L  Y  T  G  C  V  N  P
CGGCGAAAGGATTTAAACGCTTGCCAATGACC
  G  E  R  I  *
```

Potential partial preprohormone sequence:
MSAEMSKTMCLLLVVFVTVMVMGVAMAQEAPNNLNAALEYLEGLDRYYSAKARPRYGRSVGSGT
QRLYTGCVNPGERI

Signal peptide: MSAEMSKTMCLLLVVFVTVMVMGVAMA
Mature peptide: QEAPNNLNAALEYLEGLDRYYSAKARPRY
Accessory proteins: SVGSGTQRLYTGCVNPGERI

V. jacobsoni

Identified from: SRA database,

Gene structure:

TGTTTCTGTTACTCATTTTATAATAACAGCTAACCATGTCAGCCGAG**ATG**TCAAAGACGATGTG

M S K T M C

CCTTCTCCTGGTTGTTTTTGTAACGGTGATGGTCATGGGCGTGGCAATGGCTCAGGAAGCACCG

L L L V V F V T V M V M G V A M A Q E A P

AATAATCTCAACGCGGCCCTCGAGTATCTTGAAGGTCTCGACCGTTACTATTCAGCGAAGGCTC

N N L N A A L E Y L E G L D R Y Y S A K A

GACCCAGGTATGGCCGTAGCGTGGGTTCCGGGACCCAGCGTCTCTACACTGGATGCGTCAATCC

R P R Y G R S V G S G T Q R L Y T G C V N P

CGGCGAAAGGATT**TAA**ACGCTTGCCAATGACC

G E R I *

Potential partial preprohormone sequence:

MSAEMSKTMCLLLVVFVTVMVMGVAMAQEAPNNLNAALEYLEGLDRYYSAKARPRYGRSVGSGT
QRLYTGCVNPGERI

Signal peptide: MSAEMSKTMCLLLVVFVTVMVMGVAMA

Mature peptide: QEAPNNLNAALEYLEGLDRYYSAKARPRY

Accessory proteins: SVGSGTQRLYTGCVNPGERI

Appendix H – Orcokinin B

V. jacobsoni

Identified from: TSA database, GETP01194674

No match in Refseq genome database, So no prediction of gene structure

Potential Preprohormone Sequence:
MMPRHSVFALSILALSITATVWIPTVQAETNLLRREFYGPVNPELFAAFLDDHGLPREKVGQSD
RYLSRLRGGGNVARLPARSGVGKSNYDARDGQDERLSRNLDQILSPPIFVGSERETLERFGKRN
IDEIDRTAFDNFFKRNLDEIDRVGWSGFVKRLTNYLATTGHGTNTGGPVLTRRFG

Signal peptide: MMPRHSVFALSILALSITATVWIPTVQA
Mature peptide for Orcokinin B: NIDEIDRTAFDNFF and NLDEIDRVGWSGFV
Accessory proteins:
ETNLLRREFYGPVNPELFAAFLDDHGLPREKVGQSDRYLSRLRGGGNVARLP
ARSGVGKSNYDARDGQDERLSRNLDQILSPPIFVGSERETLERFG and
LTNYLATTGHGTNTGGPVLTRRFG

Appendix I – Pigment Dispersing Factor

V. destructor

Identified from: SRA database, SRR8864012.106407301.1

No match in Refseq genome database, So no prediction of gene structure

Potential partial preprohormone sequence:
KRNSELINSLLGLPKNMNNAGK

Mature peptide: NSELINSLLGLPKNMNNA-amide

V. jacobsoni

Identified from: SRA database, SRR11879882.65162670.2

No match in Refseq genome database, So no prediction of gene structure

Potential partial preprohormone sequence:
KRNSELINSLLGLPKNMNNAGK

Mature peptide: KRNSELINSLLGLPKNMNNA-amide

Appendix J – Proctolin

Identified from: Refseq genomic database, LOC111244398

V. destructor and V. jacobsoni
Gene structure:

```
GCGCCATCAGCAGTAGCGGAAGTAGTAGCCGACAACTCGAAAAACGGTGAAGCAGTTACATAACTTTATGGACGTTA
AAGTGTCTGCCTAGAAGGAGACTACGAGTGTCTGGTTTACGTCAAGTTCAACTATAAGGTGTCCACGATAGTAAACA
ACTAATACCTTTAGATTCTTCGATTATATTTATTGGCGTTGAGATCAAATAGCTTCTAGTCTCTAGAGAACAACTCA
GGATGTCGTCGCCAATGTCTGCCCGTGGCATTCTGTGGGGTCTTATCCTAGTTGGGGTGATACTACTCAGCTTGTCG
   M  S  S  P  M  S  A  R  G  I  L  W  G  L  I  L  V  G  V  I  L  L  S  L  S
GTGGAATCGCAGGCACGCTATCTTCCTACAAGAGCAGACCCGGCGCGACGAGAACGGATCCGGGAGATTCTAAGAGC
 V  E  S  Q  A  R  Y  L  P  T  R  A  D  P  A  R  R  E  R  I  R  E  I  L  R  A
 ⇓
GCTGTTACTTCTCTCGCCCGGCGAGCCCGAAGCACGCGGTCCATACACTTCGAGTTATGAATATGGCACCGGGGCGG
  L  L  L  L  S  P  G  E  P  E  A  R  G  P  Y  T  S  S  Y  E  Y  G  T  G  A
GCGATCTGAAGGGGCTTGAAAGAAGTGACTGGGATTCGTCATCGACATGATTGCCATC
G  D  L  K  G  L  E  R  S  D  W  D  S  S  S  T  *

CGTACGCAAGCTAGGGCAGTTTAAGCTGAAGCAAAATAACAACAACAGCAACAACAATAACAACAAAGCTAAGAAGA
AGCCCTAACCCTTTTTCCTGTTTGGAAGACATGTTACGAAAGAGAAGGAGAGGAAGCGGGCTACGCCGTAGCAAGAT
ATAGCAGACGAGTGGAAGAGTGGCAGAGAAAAGGAATTAAAAGCAAACGGAGGAGGAGGGGGAGCATGGGGAGGACA
AGGAGGAGATGAGAGACGAATGGAAAAAGAATAGTGCACGACACACAAAAAATGTTTCTTTGATTGCCCTTGTTCAA
                                                ⇓
GACGTTTCATGAGTCTGTTTTAGTTTTACGTTGTTGTTAGCGTTGCTTCGGGGTGTGAAGCAACTTACGCCGGACGA
CTACATTCCAGCGGAAGCCATACGAATGAATCGGCAAGTGTCCCACCACGAGTAATTGGCATCACCCAGCGCATCAA
TCGTCAAGTGGCAATCACGACCAAATTCACAATCCCCGCGAAAAGCGGTGCAGATAAAGCTTTTCAGCAGAGAAATG
AGCGCAAAAATCAATTAATCGTACATCTGATACGAGGCACGATATGCACATTATATCATGCCACACAGTACATTAAT
TAATGGAAGTAATAAATATCCGTGGATCCAGTATGTCCCATCGACCCACAGATCGAAACAGCTATCGACAGCCTATG
GAAGTGGTCCTATTAAAAAAAAAACAGCACCACTCAAATAGGCATCATGTACTCTTCTGGTACTATTACTGCTTTAC
GAGGGTCGAGTTACTATGGTTTTGAGGTTTCAGTAATGACTTCGTAGCGCCATCTTTAGGTGTTGTGGAGTTTTCCC
TCTAGGAAGGCAGAGTCACTTTAACAAGAATCAACATTCAGGCACCGCAAATATATGGCAGCACCCTCTATGGTTGT
AGGTGAAAGGTAAAATGAAACACAACTAAGATACATAATGGAGACACGTAACATCTACCCTTATTTAAAGCAAAAAA
AAACAGACAAATTATAAATAAAACTGTTTTTG
```

Potential partial preprohormone sequence:

MSSPMSARGILWGLILVGVILLSLSVESQARYLPTRADPARRERIREILRALLLLSPGEPEARG
PYTSSYEYGTGAGDLKGLERSDWDSSST

Signal peptide: MSSPMSARGILWGLILVGVILLSLSVESQA

Mature peptide: RYLPT

Accessory proteins: ADPARRERIREILRALLLLSPGEPEARGPYTSSYEYGTGAGDLKGLERSDW
DSSST